

R E P O R T R E S U M E S

ED 018 812

CG 001 271

PROGRAM EVALUATION IN MENTAL HEALTH SERVICES.

BY- PETTUS, CHARLES W.

WESTERN INTERSTATE COMMISSION FOR HIGHER EDUCATION

PUB DATE MAR 67

EDRS PRICE MF-\$0.50 HC-\$3.68 90P.

DESCRIPTORS- *CONFERENCE REPORTS, *MENTAL HEALTH PROGRAMS,
*PROGRAM EVALUATION, PROFESSIONAL EDUCATION, MENTAL HEALTH,
*PROFESSIONAL PERSONNEL,

THE PRIMARY OBJECTIVES OF THE WESTERN CONFERENCE ON THE USES OF MENTAL HEALTH DATA ARE--(1) TO PROVIDE AN EFFECTIVE CHANNEL OF COMMUNICATION BETWEEN ADMINISTRATORS AND ANALYSTS ABOUT PROGRAMS WITHIN AND BETWEEN STATES, AND (2) TO IDENTIFY TRAINING NEEDS AND ENCOURAGE EDUCATIONAL RESOURCES TO MEET THEM. THIS REPORT DESCRIBES IN DETAIL PROCEEDINGS OF THE FIRST SUB-REGIONAL WORKSHOP OF THE CONFERENCE. THE SEVEN ARTICLES INCLUDED DISCUSS THE FOLLOWING MATERIAL--(1) THE DILEMMAS, DATA, AND DECISIONS FACING DOCTORS, (2) PROGRAM EVALUATION IN MENTAL HEALTH SERVICES, (3) SOME ISSUES OF PROGRAM EVALUATION, (4) THREE SEPARATE REPORTS ON THE PROBLEMS, TASKS, AND CONCLUSIONS OF THE SIX WORKSHOP GROUPS, AND (5) IMPLICATIONS OF THE WORKSHOP. (RD)

ED0018812

PROGRAM EVALUATION IN MENTAL HEALTH SERVICES

An institute sponsored by
The Western Conference on the Uses of Mental Health Data
in cooperation with the
Oregon State Board of Control

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

The institute on "Program Evaluation in Mental Health Services" was held
October 24-25, 1966, in Portland, Oregon, and was supported in part by
Public Health Service Project Grant 1 R12 MH 09223-01 from the National
Institute of Mental Health, under the direction of Charles W. Pettus.

Western Interstate Commission for Higher Education
University East Campus Boulder, Colorado
March 1967

CG 001 271

FOREWORD

The Western Conference on the Uses of Mental Health Data is a regional association of program administrators and program analysts from both local and state level programs serving the mentally ill and mentally retarded in the thirteen western states. The primary objectives of the association are: (1) to provide an effective channel of communication between administrators and analysts about programs within and between states, and (2) to identify the training needs and encourage educational resources to meet these needs.

In addition to an annual meeting of the Western Conference on the Uses of Mental Health Data, two sub-regional meetings are planned for each year. This is a report of the first sub-regional meeting. The major states involved in this meeting were Alaska, Idaho, Oregon and Washington, although there were a few representatives from other WICHE states.

The Western Conference wishes to recognize Dr. Kenneth D. Gaver, Administrator, Mental Health Division, Oregon State Board of Control, and Mr. Calvin C. Cooper, Chief, Biometrics Section, Oregon State Board of Control, for hosting the meeting as well as for developing the agenda and securing the participation of faculty and resource persons.

Charles W. Pettus, Director
Western Conference on the Uses
of Mental Health Data

TABLE OF CONTENTS

FOREWORD	iii
DOCTORS, DILEMMAS, DATA AND DECISIONS. M. W. Neal, M.D.	1
PROGRAM EVALUATION IN MENTAL HEALTH SERVICES George Saslow, M.D.	11
SOME ISSUES OF PROGRAM EVALUATION. Jay M. Jackson, Ph.D.	27
THE PROBLEM FOR GROUPS 1 AND 2 AND GROUP REPORTS	49
THE PROBLEM FOR GROUPS 3 AND 4 AND GROUP REPORTS	57
THE PROBLEM FOR GROUPS 5 AND 6 AND GROUP REPORTS	67
IMPLICATIONS OF THE WORKSHOP E. Myles Cooper	79
ROSTER OF PARTICIPANTS	87

DOCTORS, DILEMMAS, DATA AND DECISIONS

M. W. Neal, M.D.
Chief, Office of Program Review
California Department of Mental Hygiene

Our topic today is a timely one. The advent of the new federal programs, such as the community mental health staffing and construction acts and medicare, bring hope that for the first time we can order our mental health resources into a true continuum of service. Let's not forget that these new programs grew out of dissatisfactions with the old ways of doing things and that the new ways call for new thinking.

While these new programs are still in their formative stages, we have an unparalleled opportunity to examine and revise, where necessary, our concepts of service and of the institutional organizations that deliver the services. At the same time, the advent of these new programs equally puts on us the burden of responsibility to see that these and our other state and local resources are ordered in ways that are of maximum therapeutic benefit to our patients. The resources -- manpower, money, and facilities -- are not unlimited. If they are to be used properly, they must be accompanied by information systems that report and evaluate the consequences of their use in a fashion that allows the managers of the programs at all levels -- from the hospital ward physician or the field social worker to the directors of state departments as well as legislators and others who make decisions about the system -- to make their decisions on the basis of truly timely and relevant information.

That our present data systems do not so function was clearly emphasized by the framers of the medicare legislation who deliberately singled out the

treatment of mental illness in public mental hospitals for special provisions, largely because the public mental health record and data systems were considered totally inadequate to serve as guides to audit the quality of treatment programs and the true cost of service. In part, the inadequacies of these reporting systems reflect serious deficits in personnel and equipment. But perhaps a more serious and basic flaw is the fact that many of our present mental health information systems were never intended to serve the needs of modern, dynamic, early intervention, treatment-oriented treatment programs. They have instead been developed by being grafted onto what are primarily static, vital statistics kinds of data systems inherited from a more leisurely past, when an annual census of the entries, births, deaths, and releases of a chronically static asylum population was considered sufficient. If enough bits of data are added to such a system, it may nevertheless come to have some kind of retrospective value for epidemiological and sometimes clinical research into what happened in the past. It tells very little about the present until it has become past, and it has little value in predicting the future.

It may tell us with great clarity and precision that in 1961 blue-eyed blondes from Coos Bay under 35 years of age had a higher incidence of post-partum psychosis than brown-eyed brunettes from Medford over 65, but it will have less value for ongoing management of individual patients, for monitoring the operations and evaluating the effectiveness of various programs, for helping the managers of the program to direct their resources most effectively, or for helping them detect program elements or lack of program elements which may act as blocks to patients' progress.

In general, attempts to improve the collection, retrieval, and reporting of mental health information have taken one of two paths -- either the

case registry approach or the computerized case record approach. The first aims to provide an enlarged periodically updated central data bank which may indeed be very useful for certain kinds of retrospective research. The second runs the danger of accepting the material in the case record and the systems they record without first asking hard questions about the effectiveness of the systems themselves and without careful study of who needs what information for what purpose, and when do they need it. Computerization of records without preceding analysis and improvement of the system being recorded too often implies hope that the computer can somehow manipulate the data to provide information that was never recorded by the data and this the computer cannot do.

The Western Conference on the Uses of Mental Health Data proposes you take a third path; that is, that you first study the goals and operations of representative sectors of a state's pool of mental health resources which provides a continuum of mental health services, including a state's department of mental hygiene or whatever agency is charged with supervising it, the state's local community mental health program, and all other agencies which impinge upon the mental health service pool.

Gestalt psychology furnishes the theoretical basis for this study -- that in order to understand the parts of the complex system, we must first define and understand the system as a whole. Modern management science furnishes us with the techniques for the study, such as operation research, system analysis, critical path network analysis, feasibility and cost-benefit studies, and program modeling. The electronic computer is a tool which makes our large scale studies possible. The specific goal of the study should be to develop from the analysis of the mental health system a patient career-based information system for policy planning, for case and

program management decisions and program evaluation.

In the past, this general field has not been considered the proper subject for mental health research; it was felt that funds and professional time should be reserved for more clinical kinds of subjects. We believe, however, that effective program information, and evaluation of all aspects of the program, are basic to sound clinical research and that clinical research can be valid only when there is full understanding of the total structure in which it operates, and only after artificial blocks to patient progress have been detected and eliminated. A hospital, for example, which does not convene the admission diagnostic conference until two weeks after the patient's entry will be a poor setting in which to study acute psychiatric disorder, since it is known that hospitals which are programmed and staffed to deliver early intensive care can have average total lengths of stay between 11 and 21 days.

Evaluation of a treatment procedure, which uses discharge from the hospital as one criterion of success, will be meaningless if the real determinant of discharge is inadequate social service effort to develop a family-care program, convalescent leave service, and so on. Or if the real determinant of length of stay is not response to medication but reflects failure to notify the hospital's fiscal officer to start processing the patient's account so he may have funds to tide him over the immediate hospital period. We need accurate understanding of the ways in which time is used not only to monitor research and treatment programs, but also because wasted time -- any delay in mustering services, or any unnecessary prolongation of treatment -- wastes resources.

I should, therefore, like to offer the following philosophical positions as a framework for your workshop discussions. First, that the patients

have a continuum of specific service needs that vary with a continuum of time over the patient's career. Second, that effective treatment programs are those which provide a continuum of service to match the patient's service needs. Third, that effective administration is that which marshals the program resources to match the patient's specific service needs at the place and time they will have the maximum therapeutic effect. There will be times, for example, when the presence of a social worker will be more important than a psychiatrist, when a supply of clean bed linen will be more important than an additional nurse, when a part-time housekeeper in a patient's home will be more important than medication. Fourth, that an effective information system is one that gives timely and accurate information that locates the patient in place and time, assesses his specific service needs, and feeds the managers of the system the information for decisions which allocate present service resources, predict future service needs, and evaluate the consequences. Again I remind you: by managers I mean anyone who makes a decision about a patient, whether he's a field social worker, a technician in a hospital ward, or the governor of the state.

We recognize that the total systems concepts, which have been theorized by some of the management science people such as Forrester, may be unattainable, but approaches to it must be rooted in sound understanding and reporting of the critical decision points along the career line; the time intervals between decision points; and categories of service needs. The patient's career, furthermore, is the common denominator on which all other management information systems must be based, such as accurate program cost measurements, diet planning, projections of seasonal variations in cash float to drugs, diapers, and food; in hiring and allocating personnel, and so on.

We are, let's face it, in the business of processing people. We are in the business of creating programs which help people who come in one door of our program and move through it and out the other side as smoothly and as quickly as possible in keeping with the natural history of their illness. If we are going to be in this business, we need the same kind of information to make it work that any self-respecting manufacturer of widget valves needs or he'll be out of business in six months. The sort of information he needs is what is called process information. He needs to know how many orders for raw materials are going out to his suppliers, for example. He needs to know how many box cars of steel bars are coming into the plant. He needs to know what percentage of his production is 25, 50, 75, 100 per cent completed. He needs to know how many finished goods are in his warehouse and going out to his dealers and how many orders for his raw finished goods are coming in from his dealers.

After he has accumulated this information and understands the seasonal flow and other factors that affect it, then he can program it, put it on his computer and, as long as everything is going smoothly, he can go out and play golf. If something goes wrong some place, lights flash and bells ring, then he can say to someone, "Go over and see what's wrong and do something about it." In other words he can manage his system by exception rather than by crisis, because once he has developed this kind of data and it's operating smoothly, it operates. In the same way we need the kind of a process system of information if we are going to develop programs which provide this continuum of service through which people can move.

A word I hope you are going to hear a great deal about in this workshop is the word "program." Unfortunately, when clinical people hear the word "program," they automatically think of "program" as meaning to give a course

of thorazine or electric shock, or individual psychotherapy, or group psychotherapy. This is not at all what we hope we are going to be talking about. What we're talking about is the collection and events and services that come together at the special point in time when the doctor's decision to give a pill or not to give a pill will make a real difference.

Most other medical specialties in hospitals are programmed in terms of program. When two doctors meet in a corridor of a general hospital and one says to the other, "Did you hear Joe is in the hospital?" the first question that comes back is, "Oh, what kind of patient is he?" The answer is: "He's a surgical patient, or a medical patient"; or "She's an o.b. patient"; or "He's a pediatric patient." The next question is, "For goodness sake, what's wrong?" First they think in terms of program, then in terms of the specific illness. A surgical program, for example, is not the decision to do an operation or not to do an operation or the actual performing of an operation. A surgical program is what makes it possible for a surgeon to pick up a telephone in his office and say, "Please schedule Mrs. Jones for a gall bladder operation at 10:30 a.m. next Thursday. The usual pre-operative orders," and know that at 10:30 a.m. on Thursday he can walk into the operating room, hold out his hand and have a scalpel placed in it, and begin an operation on a properly prepared patient who needs to have her gall bladder taken out. A program is what it means to have the doctor walk out of the operating room 45 minutes later, take off his gloves, and know that as long as the program proceeds smoothly, Mrs. Jones will be leaving the hospital in five days. When we're talking about program, we're talking about all of the events that may come to bear on a particular decision point including not only the clinical people but also the business administrator, the dietician, the janitor, the people who run the steam boiler

plant in the hospital, and everyone else who makes it possible for events to proceed in an orderly fashion.

In considering your workshop too, I hope that you will take to heart that we do have an opportunity for new thinking about programs. I hope that you will, for example, throw out some of the ancient shibboleths that we have in the field of mental health, such as the notion that the job of the hospital or the clinic is to cure the patient. No other part of medicine or health accepts such a dictum. Only psychiatry is stuck with this. If a doctor, for example, has a patient who has just had a heart attack, he puts the patient in the hospital to do what only the hospital is organized to do and nothing else. He puts him into the hospital so that he can hook him up to the oscilloscope over the nurse's desk which monitors his pulse and the wave pattern of his heart beat; he puts him in the hospital so that the laboratory can be there and take the blood every four hours to measure the bleeding and clotting time; so that the nurse can be only ten feet away to give him a squirt of morphine in the arm if the chest pains come back. But after the need for these things has passed, after the last test shows things are quieting down and the heart is on the mend, he sends the patient home. He doesn't believe that the hospital has to cure the patient. He tells the patient it's time to go home now. You're not well yet. You'll have to sit on the back porch for two months in a rocking chair. I'll be in to see you every day for the first week and next Monday your daughter has to bring you to the laboratory for an electrocardiogram. After two months you may be able to go to work half time and after another two months, if everything is all right, you can go back to work full time. And nobody considers the hospital program is a failure if the patient can't go out and go right back to a full-time job.

Equally I hope you will throw out the idea that a measure of a hospital's or clinic's program effectiveness is a low return rate. Return rates may in some degree measure program failure, but an effective case can equally be made that good programs may have a high return rate. With our patient with the heart attack as an example, a hospital program would not be considered a failure if a year later the man got into a fight with his boss, went out and got roaring drunk in a bar, and fell down with another heart attack. This would not be considered a hospital failure if he came back to the hospital under those circumstances. But for one reason or another in the mental health field we are stuck with the notion that the hospital and our programs cure the patient, and if anything happens to him in the future, somehow this means that our program was a failure.

PROGRAM EVALUATION IN MENTAL HEALTH SERVICES

George Saslow, M.D.
Professor and Chairman, Department of Psychiatry
University of Oregon Medical School

I set myself the assignment of trying to portray by means of a glimpse of the history of psychiatry some ways in which claims for specific and effective psychiatric treatment have come and gone. This glimpse enables innovators to see why evaluators (who tend to be rather skeptical people) have to be shown that something is any better than what's been done in the past. On the other side, evaluators need to be reminded that they generally don't innovate. They need innovators to come up with new ideas and with the enthusiasm that makes them go and become visible. Somehow persons with such different temperaments must learn to listen to each other: the enthusiast who can't bear to be stopped to answer the question "Show me that what you are doing means anything," and the evaluator who doesn't develop the enthusiasm.

I begin by reminding you that over 125 years ago a famous European psychiatrist named Esquirol, in a well-known and widely used textbook, made the following simple statement (1838): "There is no specific treatment of insanity." In the roughly 125 years since Esquirol published this statement, we have developed a few specific treatments for certain kinds of insanity. There are not many. Let me enumerate them.

The kind of thing that happens to a youngster, born deficient in thyroid activity so that he becomes a cretin, can be prevented by adequate understanding of the defect, prompt diagnosis, and treatment with thyroid extract in one form or another. Similarly we have a specific treatment for

an adult who for a variety of reasons, still not well understood, may develop thyroidism to the degree that it is called myxedema and at the same time a psychosis indistinguishable from schizophrenic psychosis. If this combination is recognized promptly and treated within less than a couple of years, there may be complete reversal of this particular kind of insanity. Then in a condition which we came to recognize as based upon specific nutritional deficiencies, pellagra, one important feature was the intellectual deterioration or dementia -- a kind of insanity. Now it is clear that treatment with particular vitamins can reverse the dementia of pellagra which often without treatment of this kind became chronic and irreversible. In the same way the kind of patient who after a long history of syphilis developed general paralysis of the insane with all its variants -- sustained anxiety, agitated depression, or megalomaniac delusions -- can now be treated effectively with penicillin. In the field of mental retardation certain kinds of treatable mental retardation have become familiar to us such as phenylketonuria. With these, early diagnosis and proper dietary alterations can prevent the mental retardation which otherwise follows. But there are not many more that we can add to the above list of treatments for insanity over 125 years.

The fact that there are not many more specific treatments for specific kinds of insanity that we can add to this list doesn't mean that many claims of specific treatments haven't appeared or disappeared over the 125 years. These claims have sometimes been for a specific treatment for a particular kind of insanity such as depression, schizophrenia or epilepsy. Sometimes there have been claims for a general treatment of insanity alleged to be highly effective for all kinds of insanity.

Few people are familiar with the extraordinary variety of these

specific claims, largely because both claims and treatments have disappeared. I thought it would be instructive to review them. Perhaps the innovators will then understand what makes the evaluators so upset when they try to work together.

Among the claims of specific treatment for one or more kinds of insanity, made since 1838, one may list the following: (1) removal of both ovaries in women, (2) removal of the uterus in women (hysterectomy), (3) castration of men, (4) trephining of the skull, (5) removing of the colon, gall bladder or other internal organs on the basis that they had some obscure focal infection which in turn caused the person to show insanity in one form or another, (6) fever therapy used for many kinds of insanity, (7) opium used for almost all kinds of insanity, (8) the rest cure of Weir Mitchell used in the last part of the 19th century for many kinds of psychiatric disabilities (9) in the 1920's and 1930's the auto-suggestion cure of the Frenchman Coué, in which you said to yourself "In every day in every way I am getting better and better," (10) the depth psychotherapies of Freud, Adler and Jung, of which only one survived. These have been claimed to have specific treatment value for all the psychoneuroses. (11) muscular relaxation, (12) hydrotherapy was at one time extremely important in psychiatric treatment, (13) hypnotism was claimed to be the best treatment available for all psychiatric conditions, (14) insulin coma as the best treatment for schizophrenia, (15) frontal lobotomy as the best treatment for schizophrenia, later for obsessive-compulsive neurosis, still later for severe anxiety, (16) electroconvulsive therapy, first put forward as the best treatment for schizophrenia and later on used mainly for something else, depressions. Then to come closer to our own time, with specific claims for achieving a great deal: (17) family therapy, (18)

conditioned reflex therapy, (19) non-directive psychotherapy, (20) group therapy, (21) psychodrama, (22) Alcoholics Anonymous, (23) Recovery, Inc. All of these have had a very interesting life history.

Before I describe what their life cycles have been in summary, I would like to read you something which you will find hard to believe unless you have read the actual sources, about the way these particular kinds of treatments which for most of us now have no meaning at all, were regarded at the particular time when they were widely used.

In 1858 in a widely distributed textbook in England, the author said: "We have learned to discriminate the conditions of mental disease in which opium becomes a true balm to the wounded spirit, a sedative in mania, a restorative in melancholia, sometimes even a tonic. The opiate treatment has gradually undergone development until at the present time the skillful and discriminating use of this drug may be truly called the sheet anchor of the alienist physician." But by 1900 this treatment was unknown.

In 1925 the famous French psychiatrist, Janet, well known for his studies on hysteria and considered to be at the same level as Freud in his imagination and ability, though working in quite different ways, wrote a book in which he discussed hypnotism. There was a period when two famous French hypnotists, Lebeau and Bernheim, felt that suggestion and hypnosis were all of psychiatric therapy, could cure everything and were all that were needed. This period lasted from 1880 to 1900 -- just 20 years roughly compared with the opium treatment which lasted 50 years. Janet said in his book, "The man (Bernheim) who was always repeating that hypnotism was everything, now himself declares that hypnotism is nothing." This is one of the few examples in which an enthusiast for an unevaluated treatment himself said "It's no good."

In 1932 a well-known American psychiatrist named Wright spoke of hydrotherapy as follows: "Hydrotherapy is the most important treatment method in psychiatry. It plays its role in delirium, excitement, agitation, insomnia (etc.), for example." There are a number of us in this audience who have seen and used large hydrotherapy installations. When I came to the University of Oregon Medical School Hospital in 1957, on the unopened psychiatric unit on the fifth floor a very large proportion -- it must have been 30-50 per cent of that very precious space -- was devoted to a brand new hydrotherapy installation. Designed in the 1950's it had wonderful marble baths, chromium-plated faucets and piping and so on for all sorts of special sitzbaths and douches coming from various countries. Not a bit of this can be found on our ward today, and none of it is missed. So between 1932 when Wright described hydrotherapy as being such an important treatment in psychiatry and 1957, not only I, but a great many of my colleagues, saw no point whatever to hydrotherapy. That was roughly a 30-year period.

In 1962 a man named Matthew Brody, who had been a member of a special commission called the Central Fact Finding Commission of the American Psychoanalytical Association and had participated in drafting a report which never was made public for various reasons, later wrote a chapter in a book in which he said the following: "The twentieth century has witnessed many advances in the treatment of emotional disorders, but it is the accomplishment of psychoanalysis that stands as the cornerstone. Psychoanalysis has emerged not only as the most effective method known for the study of the human psyche but as the most effective method known for treatment of emotional disorder." Having said this, a few sentences afterwards he says: "Our science has not yet arrived at the point where we can particularly

demonstrate its effectiveness. We have no criteria to determine degree of improvement." Since having no criteria to determine degree of improvement also characterized hydrotherapy, the use of opium, and hypnosis, it isn't hard to guess what will probably happen to psychoanalysis. I don't know how long it will take.

The life cycles of these various claims to specific treatment have varied from 20 years for hypnosis to 50 or 60 years for opium, and psychoanalysis is at the end of that range. As somebody else remarked perhaps it will end up that psychoanalysts will have had the distinction of devising a treatment that has grown lengthier and lengthier, more and more costly, serving fewer and fewer people in its original form, and with results about which we say we have no criteria for determining improvement.

Periods from 20-60 years represent the life cycle of these claims for specific treatment. These treatment methods which in general have not stood up, have had a number of important features in common. They are introduced with a tremendous amount of initial enthusiasm. They are usually identified with a particular leader, whether it be Freud with psychoanalysis, Sakel with insulin coma or Bernheim and Lebeau with hypnosis, there is always a particular leader with whom these claims have been identified. Since there is so much dissatisfaction that we have no specific treatment for most of what we call insanity, at first everybody is willing to look at the new claims. The results for the newly claimed specific treatments almost always are in the range of 90 to 100 per cent permanent and total cures. As time goes on, other persons not so enthusiastic take up the same methods. They get less impressive results. Then as more time elapses, the results become even less impressive and ultimately what has happened to all these treatments seems to have been that very limited use is found for them, or

something new is found, or the treatment is rejected totally. So as one looks over this history, one gets more and more suspicious that what is called new psychiatric treatment often seems to reflect medical, social and political fashions of the day plus dissatisfaction with our limited skill, competence and knowledge of that particular period.

I might say in passing that this is nothing restricted to psychiatry, just as Dr. Neal has made plain that psychiatry is a part of a larger tradition in medicine. Exactly the same history can be related for surgery. As late as the 1920's and early 1930's at Harvard Medical School you could find persons speaking in exactly the same way about the best way to cure tuberculous peritonitis -- it was "to open the abdomen and let in a little light and air." That is all gone now. We have demonstrated specific ways of dealing with tuberculous peritonitis. Also at that celebrated institution you could hear of colectomy for epilepsy (removing the colon) before we had adequate anticonvulsants. Operations for duodenal ulcer may take years and years to go through this life cycle when there has not been any kind of evaluation built into the original enthusiastic proposal. It also has been noted in these various other fields as well as in psychiatry that the best results are always obtained by the initial enthusiast. The enthusiast for a given kind of surgical procedure always gets better results than the skeptics, whether the results are evaluated objectively or subjectively.

I want now to go into two recent developments which concern all of you who are doing clinic psychiatric work and who are in a current innovative phase. There are two recent developments which are very striking in contemporary psychiatry. One has to do with the great hopes that were aroused by the new drugs that began to be used about 1951 when chlorpromazine was

introduced. The other one has to do with community psychiatry. These are two of the dominant themes at the present time.

With regard to the drugs it has already become clear that they will be no magic cure for insanity. Their role has rapidly become much better understood in a much shorter time than the usual 20 to 60 years that I have described for non-evaluated new claims for specific treatment. At the same time a very interesting feature of this situation I think ought to be in your awareness. So many drugs can be made per year by modern organic chemists that it is impossible for them to be properly evaluated as they appear. Doctors have these drugs urged upon them by the pharmaceutical companies which stand to make a great financial profit from the widespread prescription of any new drug which is successful. The initiative in evaluating drugs, the initiative in pushing them has apparently been lost by the clinical practitioner, by the university research department and by any other interested research group. Control of the introduction of this particular kind of claim to better specific treatment is now mainly in the hands of the commercial companies. It is interesting that when not much money can be made from some kind of preparation such as the very cheap lithium salts for the treatment of mania about which there have been some very encouraging reports from Great Britain, you don't observe these being pressed upon you in your office or your hospital by anybody. One feature of new claims for specific treatment involving drugs has to do with the fact that we are not in control, that is the professional people are not deciding which drugs to use, whether we need all of the new drugs, and which one is better than which. I know of only two nongovernmental non-commercial sources of information about effective drugs. One is called the Medical Letter, edited entirely by doctors, and the other one is the

International Drug Therapy News. These present the reports of well-designed tests on the effectiveness of this drug or that by comparison with placebo or some other drug about which we already know something. It is in such publications that you will find that drug holidays have been demonstrated to be safe for chronic psychotic patients. Some may be taken off drugs altogether, some may be on drugs 5 days out of 7, or 1 week out of 4, etc. Such intermittent use of drugs has been shown to be as satisfactory as continuous drug administration, for many patients. Continuous drug intake is a very expensive proposition for all psychiatric installations and patients.

Now to come to the community psychiatry interest which is now so dominant. At the present time community psychiatry can be viewed as a social action movement among other things. It is influenced by public pressures of great magnitude, by political forces, and by legislative measures at all levels -- local, state and federal. Enthusiastic administrators dissatisfied with how we are doing things now, optimistic laymen, enthusiastic clinicians all participate in pushing us in this direction.

Community psychiatry has at its theoretical basis the social-behavioral sciences rather than the very narrow biological model of American medicine. I might say in passing that it is not widely known that in European countries social responsibility in all branches of medicine, even at the national level, and prevention as well as comprehensiveness and continuity of care, have long been a part of the system of health care. Their views of man as a biological organism have never neglected the psychological, the interpersonal, the social organization dimension to the degree that we have apparently fallen into.

This distinction, then, between a narrow biological medical model of health care and a social-behavioral science model, is not as valid for

European countries as it is for us. Community psychiatry has its theoretical basis in the social behavioral sciences rather than in the narrow biological model of medicine focussed on organs and biochemical and physiological processes. Community psychiatry broadens the definition of the psychiatrist's job to include social problems never before included, such as alcoholism, sex offenses, delinquency, drug abuse, family disorganization. In addition, community psychiatry includes certain areas which many of us have thought a proper concern of psychiatry but which have been neglected -- mental retardation, chronic psychosis, senile deterioration. Community psychiatry emphasizes the role of the social unit in which the patient functions as a maintainer of his self-defeating behavior and as a reinforcer of more effective behavior. This social unit may be the patient's family, his ward, his school, his work setting or his community. The therapist who takes this point of view now has re-defined his own role to be that of a community leader and educator in considerable part as he collaborates with many other types of professionals and as he considers treatment in the community, in the half-way house, in psychiatric units in general hospitals, in schools, in clinics, in day hospitals, and in all of the various ways in which treatment is now discussed under the heading of community psychiatry.

How do we determine which of these various activities has therapeutic results that can be confirmed? That is a tough question. You will find that Dr. Neal aided by computers is very anxious to wrestle with it but he needs your participation. Such evaluators can't work on you and observe you from the outside as relative non-participants and begin to answer such a question as: How do we determine which of all these activities has therapeutic value? You have to be participating with the evaluators. I

pointed out that this is not easy for innovators. Innovators don't want to be bothered with questions such as does the new treatment really make any difference in the outcome, and evaluators are bothered with nothing else so much as does it make any difference. Somehow you are both going to have to talk together, to raise the appropriate questions about the issues which have to be examined and to find scientifically acceptable ways of doing it.

We have learned certain things from the history of psychiatry which I have described in terms of claims for specific treatments which have come and gone. We have learned that certain things are not going to help us decide which of various therapeutic activities that are new and arouse our interest are worthwhile and will last. It is a sad list. Plausibility, for example, has turned out to be a very poor guide. We have had very able professionals in the field come up with the most plausible reasons why you ought to remove a colon for epilepsy but it hasn't meant a thing. Plausibility then has turned out to be a very poor guide by itself. Humanitarianism has turned out to be a poor guide. The most extraordinary things have been done to do people good but haven't got them well or less sick. Enthusiasm has turned out to be a very poor guide. An inspiring leader has turned out to be a very poor guide. Optimism has turned out to be a very poor guide. It is clear that a lot of things on which people tend to rely are not going to help us deal with such thorny issues as whether the millions of dollars that are being poured into the contemporary innovations in psychiatry really will turn out 25 years from now to be very much more justified than the high hopes we had for the prevention of adult disorganization when we set up the first child guidance clinics in 1922 in this country.

It looks then as if what we are going to have to do is deal with the

question of what is therapeutically viable, what can be confirmed about our guesses so far as innovative approaches go. It looks as if what we are going to have to rely on will be the same exacting techniques upon which we learned to rely in other branches of science: the same powerful scientific methods, not very exciting in some ways, which have led to such discoveries as niacin for pellagra, thyroid extract for cretinism, penicillin for syphilis. None of them involved tremendous efforts costing millions of dollars. Many claims can be tested by appropriate procedure in a pilot project, by proper questions, proper controls, and relatively small amounts of money.

Have well-tested scientific methods ever been used in any field as complicated as the one we're dealing with? The answer is that such techniques have been used recently in matters that approximate in some ways those with which psychiatry has to deal. I'll mention one striking example and allude to a second before I conclude.

Some years ago, about 1956, the news spread abroad among surgeons that a way had been found to increase the blood supply to a failing heart in people who have many attacks of angina pectoris. Some people have these many times a day. Their origin is thought to be in an inadequate blood supply to the heart because of reduced blood flow through the coronary arteries on account of coronary artery sclerosis which we don't know how to prevent. Among the ideas which were being talked about at this time was increasing the blood supply to the heart by having new vessels grow into the heart. This, it was claimed, could be done by a very simple operation which could be done under local anesthetic involving two arteries that run right along the sternum called the internal mammary arteries. These arteries can be exposed under local anesthetic, one on each side of the sternum. One can slip a ligature around the artery and tie that. By

blocking blood flow into the internal mammary arteries, blood flow into other arteries that could carry blood to the heart muscle would be increased. The most enthusiastic reports were published by a few surgeons about the way in which the number of anginal attacks after this operation decreased almost to zero, the way in which the number of nitroglycerin tablets people had to take to relieve the unbearable pain decreased markedly, and how these persons could carry out standard exercises for a noticeably longer time without pain than before the operation. Skeptics raised the question whether there had been adequate scientific design of the studies so that the new operation could be considered of demonstrated effectiveness. Over the next two years (1956 to 1958) in three well-designed experiments in two widely separated cities the following was done. Patients (36 who suffered from severe anginal attacks) were told that it was quite uncertain how long the results would last. It had already been learned that the first enthusiasm of doctors and patients hadn't led to sustained improvement past five or six months. Patients, therefore, could quite honestly be told that nobody knew how well the operation would work, and be invited to help the surgeons evaluate the results very carefully. Patients were selected for this operation on the basis of their need of a better blood supply to the heart. At the moment that the surgeon was to operate on the patient he was given a card which came from a table of random numbers and the card said either "operate" or "sham operate." Sham operate meant that everything was done exactly as had been described with such enthusiasm in 1956 except that the two arteries were exposed, a piece of absorbable ligature was slipped under each one and it was left tied loosely and covered over. The arteries weren't tied off tightly at all. That was the sham operation. In the actual operation the arteries were closed tight by ligatures. Then over the

next period of several months the various standard observations were made, how the person felt, how many attacks of angina were recorded, how many nitroglycerin tablets they took, how much exercise could be borne before pain was unbearable, and before certain standard electrocardiographic changes appeared. The results of these three experiments were the same. No difference in outcome could be demonstrated between the sham operation and the actual operation. So by 1958 -- a period of only two years -- this operation no longer seemed justified to surgeons. It took just two years to settle the question.

The second example is on a much larger scale. The asking of appropriate questions and the setting up of procedures designed to get answers about the effectiveness of innovative educational approaches is actually going on in the field of secondary school education and has been since about 1954. In field after field, starting with physics, then mathematics, then biology, chemistry, and ultimately English, history, music, etc., a carefully thought-out approach like that used in the internal mammary artery studies has been adopted. Each innovation is being tested by appropriately designed comparison procedures. In this country it has taken roughly 40 years for a new educational procedure to get in and for an old one to vanish, often without any evaluation of either. Millions and millions of dollars are wasted in the process. Since 1954, by raising questions in this particular way and by having innovators and practitioners, that is the enthusiastic teachers who are dissatisfied with past procedures, the administrators, and the research workers on the frontier of mathematics, on the frontier of physics, on the frontier of how to teach people to enjoy literature, these have all learned to communicate with each other and with competent evaluators. No innovation is introduced in these revolutionary new curricula

without a simultaneous evaluative scheme. These evaluative schemes have already turned up the most important information. Some of the innovative curricula, for example, which often reduce the hours spent with a teacher, are even more beneficial for the poor learner than for the supposedly very superior learners. All sorts of new abilities have turned up in youngsters as these innovations have been tried. Inarticulate youngsters doing badly by conventional school criteria have turned out often to have the best suggestions for ways of finding out things in the special settings created in the new curricula. When they work with a group their mechanical ability can be used, their ingenuity can be used -- abilities which were masked previously by the conventional criteria for educational achievement.

It has cost five million dollars to get one of these new curricula going, to involve 20 to 30 per cent of the high school students in the United States that are in such a field, to develop the new textbooks, to develop a new way of instruction in the field. Compare this cost with the educational bill in our country which was recently twenty-four billion dollars a year. How do you know whether the innovation which to you seems so plausible and about which you are so enthusiastic really does anything that you yourself would consider worthwhile X years from now? Such questions have led to many extraordinary changes in high school curricula, so that we are helping young people learn things which many of us learned only when 10 or 12 years older, if even then.

It is clear then that we're going to have to do thinking like this somehow, that it is very worthwhile to do it, and that innovators and evaluators in our field don't know yet how to do it together. As a beginning they are going to have to ask such questions as: What are the issues about which we should be doing evaluating and working together in relation

to our current enthusiasm? It is this subject -- what are the issues about which evaluative questions need to be raised? -- that has been left for Professor Jackson this afternoon. Our brief glimpse at the history of psychiatry tells us what will likely happen to most of what we are now convinced is worthwhile as treatment if we don't direct our attention to such issues.

SOME ISSUES OF PROGRAM EVALUATION

Jay M. Jackson, Ph.D.
Professor of Sociology
University of Oregon

The purpose of this paper, as I understand it, is to raise issues which may stimulate your subsequent discussion today. Anyone knows that it is much easier to ask questions than to answer them, and I intend to take full advantage of this opportunity to point to issues in the evaluation of programs for which I have no solution. Where I do have some proposals or suggestions that may be worthy of your consideration, I shall feel free, naturally, to advance them with great vigor.

Although there are a great many complex and difficult problems involved in the evaluation of programs, some of which have been discussed already with clarity by your previous speakers, I wish to address my remarks to three basic questions, and to the issues and implications that arise from them. First, there is the question that, of necessity, must take priority over all others: why should evaluation occur? Second, it will be important to consider what it is that needs to be evaluated? And finally, we must ask ourselves who it is that should be doing the evaluation?

Although we are gathered here to discuss the evaluation of mental health programs, it appears to me that the issues involved are not specific to the field of mental health and mental illness. I would hope, therefore, that my remarks would have some general applicability to all types of program evaluation.

I. Why Evaluate Programs?

The first question that must be asked, in discussing the evaluation of

programs, is "why evaluate?" Let us begin by challenging the basic assumption that evaluation is always necessary or desirable. After all, there are many kinds of activities in society that do not get evaluated. In fact, we would have to say, all of the important ones. We certainly do not really evaluate education, at any level. We just believe that it is a good thing. We do not evaluate religion. We do not evaluate the family and raise questions as to whether we should or should not have one.

Since most of the basic activities in which we engage are not evaluated, but depend upon faith in our values, it seems to me that we should not feel guilty at not having the answers, at not knowing how to evaluate our day-to-day endeavors in the area of mental health. For one of the things that I am going to say is that, for most of the programs that we are attempting to evaluate, we are not going to be able to do it, at least in the near future, at least in the way that we are trying.

In considering why programs should be evaluated, it is important to distinguish between questions of value and questions of fact. It certainly is possible to discuss this issue in terms of should or ought questions. The speakers this morning, for example, were on a very high moral level. They were talking, implicitly at least, about the ultimate goals of science and the good of all mankind.

Knowing a little about human nature, and knowing something about our own and others' natures, let's face some facts. We are not always concerned with evaluating programs in order necessarily to find out whether they are as good as they ought to be or could be. Let us get down to realities rather than discuss ideal values. Given a concrete situation where we are involved in some attempts to evaluate a program, what are some of the real reasons for attempting to evaluate it, as you and I have experienced them?

Granted, that we are all seekers after truth, are there not a number of other, somewhat more mundane reasons for evaluating a program?

For example, the terms of the grant that we have accepted may have included the requirement that we evaluate what we are doing. In order to obtain the opportunity and the resources to accomplish something that we are convinced is socially valuable, we accept this condition. Thus, the pressures to evaluate the program do not arise from the intrinsic needs of the operation, but from the external environment. There are forces from elsewhere demanding that we do some evaluation, whether we want to or not. Is this not one general reason for undertaking the evaluation of a program, that this is a built-in condition of the program, specified by a particular set of others who may be distant from the actual operations in time, space, and knowledge?

Thus, taking a factual approach to the question, we can recognize pressures to evaluate a program which do not arise from its intrinsic requirements. Without questioning whether they are good or bad, we can recognize that they do exist.

There are other such reasons for evaluating programs. Suppose, for example, that we have a program which is encrusted and enfeebled by the ravages of time and effort. The personnel are weary or cynical, and have really not had a new idea or a new activity in years. You have all seen programs like that around the country. Evaluation of such a program may be used as a lever, to try to shove people out of the situation, or to move a program when everyone is convinced that it is doing nothing. And if it's doing nothing, it cannot possibly be doing anything very useful!

Sometimes we evaluate a program because we are convinced that it is worthwhile. We are looking for some type of evidence that will help to

convince people to continue to support, or to expand this social enterprise. Under such circumstances, it may be that we are not always solely concerned with the basic scientific goal of discovering the truth. Evaluations of programs often are utilized as ways of influencing people, for public relations, or for generating social action.

Why is it important, if we are not engaged in moral indignation, to penetrate beneath the facade and to confront the real reasons for evaluating a program? I think that it is important, because unless those who are involved in the evaluation of a program are quite clear and in agreement about why they are doing it, they are likely to find themselves in great difficulties. They will probably be engaging in misperceptions, miscommunication, and, in general, confusing and frustrating one another. If one person or one group thinks that the objective of the evaluation is to "find out," and another person or group thinks that the objective is to "convince," then they are going to have problems!

A somewhat different way of approaching this question is to recognize that there are two main types of goals for evaluation, or the results of evaluation of programs. One goal can be called absolute. We wish to determine the scientific truth about a program, exactly what effects it has under given conditions, within a specified degree of probability. For a great many reasons, some of which have been suggested already in this paper, some of which will be discussed later, I think that this is an unrealistic objective. Another type of absolute goal is to determine just how good or bad, how valuable or useless the program is. This may sound very much like the preceding one, but it is different in that its objective is criticism and the allocation of praise or blame, rather than the generation of knowledge. This is an ethical rather than an epistemological quest! Very often the

results of such an evaluation may be phrased in terms of certain people doing a good job and certain others doing a poor job. This is usually a useless result of an evaluation. What are you going to do with it, especially if those are the people that are available to the program? And what do we mean by good or poor? Compared to what?

I do not believe that any program evaluation can in fact come up with a conclusion that is not relative and comparative, even though we formulate the objectives in absolute terms. Any scale that we use, any units that we count, any measures of the results of a program, are going to be expressed in numerical form. And these numbers have meaning only when they are located within some distribution. This is what permits us to determine that they are relatively high or low on the scale. You all must recognize that this is basic to elementary measurement theory. But its implications are often ignored -- that the evaluation of any single program is meaningless by itself. It must be evaluated in comparison with identical or highly similar programs, utilizing the same measures.

Since programs tend to be complex, and to differ with regard to at least their geographical location, their personnel, and their clientele, it would appear that each program is unique, and that accepting the dictum that evaluation must be comparative would lead to the conclusion that it is not possible to engage in meaningful evaluation of programs. I believe, however, that programs are unique only in the same way that individual persons are unique; they become similar and comparable to the degree that we discover common or universal dimensions that can be measured under relatively standard conditions.

One of these standard conditions is that programs have comparable inputs. There does not seem to be much point, for example, in comparing the

results of one hospital which spends \$10.00 a day per patient with another which spends \$5.00 per day. Comparisons between treatment institutions with highly different facilities, or quality of personnel, do not lead to meaningful evaluations. As part of a controlled experimental design they might conceivably tell us something about the effects of facilities or personnel upon outcomes; but this is a different question which should not be confused with program evaluation. The basic principle here seems to be that comparisons should be made between programs that are similar in their resources and their conditions of operation.

This line of reasoning is based upon the assumption that no evaluation is any good unless it has implications for improvement or change. It's not enough to evaluate something and say that it is good or poor, low or high on some scale, or even that compared to other states that it is better or poorer. So what? When your evaluation has been completed, what actions are required? And what actions are possible?

This last question needs to be considered at the outset of any evaluation of a program. What is the range of tolerable results of the evaluation? It may well be, for example, that the system in which the program is embedded has no tolerance for a result which says, in effect, "the program is no good, scrap it!" There may be too many commitments, too many personnel, and too many reputations involved. In which case, it does not make much sense to come up with that kind of answer. It may well be that the range within which change in a program can occur is relatively slight at times. This needs to be determined in advance, if possible, so that the objectives of an evaluation can be geared to the tolerance that a system has for change.

Assuming that change in the program is feasible, and the evaluation has

led to critical conclusions, it is clear that the results must be detailed with respect to the various components of the program. What you want to know is in which respect the various sub-parts of the operation relate to the outcomes. It certainly is not enough to know that a total program has a certain degree of over-all effectiveness. What are the parts of it that are contributing to the assessment, and how can each part be changed or improved? An evaluation which does not pin-point the aspects of a program that require change or improvement is not doing its job.

II. What is Evaluated?

We have been discussing programs and their evaluation, and will be using the word "program" quite frequently during this conference. We seem to use the term in many different ways. I have often wondered when a program is a program. People write to the university and they ask about a specific graduate program, and I think, well, do we have a graduate program? We have some courses and some students and some instructors, we have some rules and regulations, but do we have a program?

If something is ongoing, like a mental hospital, for example, (even though we hope some day to bring them to an end, still, they seem to go on indefinitely) is this a program? Or does a program have to have some limited time duration, a beginning and an ending? Like this conference, for example. Is this a program? It has a beginning and an end; it has inputs and a goal; it has some resources and some activities. Do we know what we mean by a program? Is an organization like a home for the mentally retarded a program? Or do you think of some part of it, some discrete set of activities, as a program?

This is not an easy question to answer, but if we do not know what we mean by "program," how can we be sure what we mean when we talk about the

evaluation of programs? It seems to me we ought to be able to define what we are talking about.

We heard something about programs this morning from Dr. Neal. I believe that management science people are very clear about what they mean by a program. They talk about a program as some set of activities -- what I would call a social enterprise -- that has certain inputs of resources and conditions, certain ways of organizing these resources and conditions and establishing relationships among them, and certain outputs with standards for evaluating them.

This seems to be perfectly straightforward, but one of the difficulties with most of the mental health programs that do get evaluated -- I am thinking of demonstration projects supported by NIMH, for example -- is that the goals are quite vague. Let me illustrate by quoting one of them: "Something can be done to assist the patient to function more adequately in the social world and thereby to be less drain on the energies of those around him."

Now, suppose that this is one of the objectives of a program. In evaluating the degree to which it is achieving its goal, in trying to develop some system of measurement to quantify the program's effectiveness, you have to decide what "function more adequately in his social world" means. How do you measure this? Can you measure it before you conceptualize it? Surely, you cannot measure anything very well unless you have a pretty clear idea of what it is. It seems to me that the better your ability to conceptualize something, the better your ability to measure it.

What does it mean to "be less drain on the energies of those around him"? How do you go about operationalizing it, obtaining measures of it? Clearly, this is going to be a difficult outcome to evaluate, but I believe it can be done.

What we do is decide arbitrarily on definitions of our terms and make them very explicit. Our evaluation will then stand or fall depending upon whether other qualified people agree with our definitions. If we can find operations for "adequate functioning in the social world" that enough people will agree with, then we can proceed to evaluate the program using those criteria. There will be enough colleagues and people in the field who will be willing to accept our evaluation. Not everybody, but if we wait until everybody agrees with our operational definitions, our criteria, we shall wait forever. Evaluation certainly requires that we stick our neck out and decide quite arbitrarily what we mean by certain terms, even though that is not all of the meaning. Even though we do not capture the whole of an idea, at least we capture some of it.

One of the great difficulties in program evaluation is that people utilize different definitions. Very often we do not talk about these. It seems kind of bush among professionals to discuss what we mean by mental health, or what we mean by adequate social functioning. Most of the things that we take for granted are relatively undefined. They represent some kind of ultimate faith. One of the issues is, should we go ahead and leave them undefined? Should we simply have faith that we are doing something constructive and useful, or should we spend our time talking about the meaning of some of these hinge concepts, these hinge ideas upon which so much else depends?

Another reason why the goals of mental health programs are often vague is that there are different types of goals which frequently go unrecognized. This is pointed out in a very good and useful book that some of you may know about called Psychiatric Rehabilitation: Some Problems of Research by Kandel and Williams. They make a distinction between primary prevention,

secondary prevention, and tertiary prevention, three terms from public health, in relation to the rehabilitation problems of mental illness. Primary prevention is reduction in the incidence of disease; secondary prevention is the early recognition and tentative treatment of disease; and tertiary prevention is the prevention of disability after disease.

Thinking of these as possible goals of programs, you can see that often we are not clear as to what we are really trying to do. Are we trying to identify and to treat cases as quickly and definitively as possible, or are we trying to prevent people from lying around mental hospitals becoming more and more institutionalized? In other words, disability after illness. The question of clarifying goals certainly is very important, before an evaluator can proceed to select criteria for his evaluation. However, that raises some real problems.

Mental health programs, especially those which incorporate current therapeutic community, total push, and related treatment philosophies which pay attention to a patient's total environment, including the treatment institution, the community, and the family, are highly complex and becoming more so. The examples of evaluation that we had this morning, even the complicated heart surgery case discussed by Dr. Saslow, were basically simple compared to the kinds of programs we wish to evaluate. Here I would like to read to you another example from Kandel and Williams. This book, incidentally, was the product of a conference in which 49 different projects, each in the area of rehabilitation of the mentally ill, were discussed. Most of them involved attempting to evaluate a program. Let me quote a description of one of these programs:

"One project established a special facility to provide group living for post-hospital mental patients who were at the same time being assisted

in vocational rehabilitation by a local employer's rehabilitation planning committee. A house was rented in a residential neighborhood and made into a residence for male patients. It was run on a self-governing basis with the help of a housemother, a resident supervisor, and a psychologist. Vocational and individual supportive counseling on a 24-hour basis to help in securing a job were provided to each resident by the supervisor. Patients with severe psychiatric problems were referred to appropriate agencies...." And so on. There is quite a bit more of this description, but I feel that I have read enough.

Suppose that you are trying to evaluate this program. You are using an input-output model, as most evaluations do, which specifies some type of input into the program, some intervening processes, and the output or effects of the program. Suppose that you obtain results which demonstrate that the program is doing an effective job, compared to your control group or to some other experimental situation. Do you know, in fact, which part of the program is responsible for your results?

How many inputs are there into the program I just described? There are a large number of activities, each dependent upon certain resources, human, material, and social, and many environmental conditions. I suppose that the number of inputs is just dependent upon your ability to conceptualize them. This example seems to me to be much more typical of the programs that we wish to evaluate than the laboratory experiments which implicitly or explicitly provide our evaluation models.

Even when you do an experiment in a laboratory with white rats, it is much harder than the textbooks say to achieve adequate control groups, adequate observation conditions, adequate insulation from environment. If any of you have had the experience of conducting experiments on people in

a laboratory, you know just how imprecise these are. But these are still in the artificial situation of a laboratory, where you are attempting to vary one or two variables at a time and to control the others.

In our discussions of program evaluation, we are confronted with complex social situations which have literally hundreds of things occurring at the same time. The problem in such situations is to know what part of the total complexity is related to your outcome measures. We can talk all we like about changing one thing at a time and keeping everything else constant but this is impossible. It just does not happen.

The simpler the act that you are trying to evaluate, the more control you can have over it. But something, even to be considered a program, must ipso facto be complex. Certainly, the types of programs which we wish to evaluate are highly complex. It seems to me, therefore, that the traditional input-output model really is quite useless.

The continuous information generating model that Dr. Neal was discussing this morning is certainly a better model. The basic problem here is how do we generate the important kinds of data? We find that the kinds of data which are easy to obtain are usually less useful or more superficial. Very often, although certainly not invariably, counting the number and kinds of patients who enter into the various stages of a program, or emerge from them -- the global kinds of data that are readily accessible -- do not really tell us what we need to know. It is much more difficult, however, to get into the intervening processes of a program.

One of our great problems here is that most people working in the area of mental health and mental illness have been provided with an individual model in their training. They think in terms of individual persons, not in terms of systems. After all, what would you expect? They are concerned

about mental illness which is, presumably, a "disease" of an individual. They are confronted with individual cases, with persons. They are concerned about persons; and so as psychiatrists or psychologists or social workers, they think about individuals. However, when it comes to the evaluation of a program, it should be realized that the program is some type of system. Any type of social enterprise, conceived as a system, leaves most mental health workers without adequate conceptual tools for thinking about it. How do you conceptualize a system? What kinds of concepts do you use?

One of the solutions often attempted is to extrapolate an individual model to an organization. I remember when the industrial research group of the Menninger Foundation began doing studies of organizations. It seemed that organizations could have all types of repressions and unconscious motives. I do not know whether they can suffer from Oedipal complexes or not! It was quite clear that the whole conceptual terminology had been extrapolated to the level of organizations. In the absence of any other conceptual scheme, we have to use analogies. We use whatever conceptual tools are available to help us in our thinking. But sometimes they are quite misleading. After all, organizations are not persons.

In reviewing applications and proposals to study mental hospitals, one often perceives a tendency to use the term "social system." This is kind of fashionable these days! I think that it improves the probability of having an application accepted -- at least, in the minds of the applicants -- if there is a generous sprinkling around of social system concepts. When one reviews such applications, it becomes pretty apparent that these concepts are not doing anything for the proposal simply because they are not being utilized appropriately.

It was enlightening to examine the kinds of problems that the staffs of the 49 projects in the Kandel and Williams book had, in attempting to conduct their evaluations. It is ironic: here we have some people building a project to study a program, without realizing that they, too, have built a program. (You can get, if you like, an infinite regress. Now you really need somebody to study the program that these people have built to do the evaluation of a program.) It is quite clear that in each project they had a program going themselves, the objective of which was to make an evaluation. The kinds of problems they report in Kandel and Williams are difficulties in their own program: problems of communication, the role system within their project -- many different types of people were quite unclear about what their roles were and how they related to one another. One of the biggest areas of problems concerned the authority structure. It was not always clear who had the right to make what decisions about what aspects of the evaluation project. In concentrating on the program that they were evaluating, they rarely looked at these kinds of problems, because they themselves were not trained to conceptualize a program as a social system. They did not think of it as a system of roles, as having an authority structure, a communication structure, and so on.

I believe that this is one of the basic requirements of people who are trying to evaluate programs. They need a set of conceptual tools to view it as an interlocking system rather than as a lot of discrete individuals with their personality problems, fitting into various kinds of typologies. Often, this is how a program is explained to be effective or not. Personality concepts are primarily used in making evaluations; in the worst cases, the same concepts are used for both staff and patients. It raises the issue, when we are evaluating a program, are we evaluating the personnel,

or are we evaluating the program?

This question has some important implications for yet another issue: what effect does the evaluation have on the system? There are two main approaches to answering this question. You can approach it from the viewpoint of pure science, which says that since we wish to obtain the truth about the functioning of the system, we must not interfere with it. This is a basic problem of all science, to what degree does the study or measurement of a phenomenon affect it?

Given that in the evaluation of a program we are not doing basic research, but are concerned with understanding an ongoing operation which is having many human and social effects, there is another question which I think is equally or even more relevant. Is there any danger that by evaluating a program we are diminishing its effectiveness? There is no general answer, of course. We have to specify under what conditions there will be a greater or lesser probability that an evaluation of a program will harm it. You can begin to speculate about this. Clearly, it depends on who is doing the evaluating for what purpose. This repeats one of the issues which I raised previously. Why evaluate? It also raises the issue, who should be doing the evaluation?

III. Who Should Evaluate?

Listening to my distinguished colleagues this morning, I was struck by the fact that neither of them gave us a picture of evaluation going on within a social structure. Dr. Saslow told us about various experimental studies that had been conducted throughout the history of medicine and psychiatry. It appeared as if each of these experiments had been conducted in a vacuum. Implicit in the entire discussion was the laboratory model. We have previously suggested that this model is not appropriate for evaluation

of complex programs. We have here yet another reason why it is not adequate.

All of us have bosses or superiors, and probably all or most of us have subordinates or employees. We work and live in socially structured situations. These relationships we have with other people, with other positions and groups and professions, affect how we do our jobs, how motivated we are, and what we think of ourselves. We all know this. Yet nothing was said this morning about authority relations, which I take to be one of the more critical areas in the evaluation of any program. (This is something we rather seldom examine, because sometimes that's us!)

The question, who should do the evaluation, is very much related to what effect evaluation is going to have upon the system. This, of course, is dependent upon the kinds of relationships members of the system have with one another. To what degree do they threaten one another? To what degree are they competing for scarce rewards? To what degree are they interdependent and cooperative? All such questions are highly relevant to who should do the evaluation.

If you are going to conduct an evaluation which studies the program as a system -- which means getting into the innards of the program, and seeing the ways some parts are related to other parts -- then you have to begin by describing and understanding such relationships.

The structure and type of relationships among the staff of a program are as significant for the outcomes of its activities as any other factors. After all, when you think of a mental health or a rehabilitation program, what are the inputs? You might say money, but what does money buy? It buys staff, it buys people. These are the resources of our programs, by and large. We are working with people, and these people in the mental health field are as dedicated and motivated as any group of people you can find in

any human enterprise -- at least when they enter the field. Otherwise they would not enter.

The big problem in many of the programs that I have seen is the tremendous waste of human intelligence and human energy. Over and over again I have talked to people who tell me the things they would like to do. They say that they are not satisfied with what they are doing. I ask, "Why don't you change?" They reply, "That isn't the way things are done here. It isn't the system."

It appears that the "system" qua system is bigger than any individual or several individuals. Then it is the system which we need to study. The system involves the authority structure, the communication structure, the allocation of rewards and privileges among members, and other such relationships which we seldom study or discuss.

Typically, the evaluation of a program is done by those near or at the top of the status pyramid in an organization. The people on top often are in the position of evaluating the efforts of those on the firing line. It is necessary to ask how these subordinates will react to the evaluation. Since many of the most important data we require are forthcoming from these people at lower echelons, and they are the only ones who can provide this information, it is rather important that they not be afraid or threatened by the evaluation. If they are, then are we going to be told the kinds of things that we want to know, or are we going to get the kind of data that they think that we should have?

One way of attempting to avoid the distortions of data that depend upon what subordinates are willing to tell superiors, is to rely upon observation. Investigators often believe that they get around this problem by observing behavior directly. They do not interview people, they do not

ask them questions, they just observe them. Anyone who has tried this, however, knows that behavior changes under observation, depending upon the assumed purposes of the investigation and the relations between the investigator and the subjects. The problem, of course, is especially acute when an evaluation project is being conducted by those at the top of the authority structure.

It is often not possible to determine when behavior is being changed. I have had the experience of entering a large, ongoing operation under the auspices of top management, without any announcement or fuss, just to observe what I could. I thought that I was being ignored by the workers in this situation, all busily going about their business. It was not until much later, when I really got into the system, and got people to trust me, that I learned what had gone on the first day that I arrived. Different people were asking, "Who's that, and what's he doing here?" There was a tremendous amount of communication traveling around the grapevine. People are very much concerned when their behavior is being observed, and rightly so, when it is being evaluated.

One must conclude that evaluation by the top people in a system is unlikely to produce any objective or definitive understanding of how and why a program achieves the effects that it does. This is so because of the inaccessibility and distortions of data, and also because people at the top are unwilling and unable to study their own activities and influences, which are often the most significant aspect of the program.

Another approach, one that is frequently tried, is to bring someone in from outside to do the evaluation. Here again we find that the kinds of data that you really want are rarely accessible to an outsider, without a long period of building acceptance in the system. Dr. Jim Bosch, one of

our conference participants, knows as an anthropologist the difficulty that he and his colleagues have in entering a new situation and seeing anything that is really important for the first six months or a year. You just are not permitted to get in to see anything. In brief, the problem of having an outsider come in to do the evaluation is that he very seldom can learn the things that are really important.

Then, who should evaluate a program? In my concluding remarks I want to raise one more thorny issue, with a suggested approach. Dr. Saslow this morning proposed that one of our real difficulties is that there are two types of people, what he has called innovators and evaluators. I am not sure what he means by innovators, but I suppose that anybody doing an activity in a program is an innovator, since he never does things exactly the same. In other words, he was referring to the people engaged in an activity, who want to get something worthwhile accomplished, and the people evaluating it, who want to be certain that it is indeed worthwhile and efficient. I believe that there is some truth in the proposal that there are temperamental differences involved in these two different types of role occupants, although I cannot accept the dichotomy as between people. It is more probable that differences in the demands of the two roles account for the divergent concerns and behaviors of those engaged in an activity and those attempting to evaluate it.

Perhaps the two roles ought to be combined into one, and the people to conduct the evaluation should be those performing the activity. I think that a strong case can be made for people evaluating their own activities. Not in terms of an input-output model, since we have seen some of the shortcoming there, but more in terms of a continuous feedback process. Assume that our programs are as complicated as I have suggested they are --

highly complex. Assume that we can never really definitively attribute the outcomes to the influences of any particular sub-part of a program. It becomes apparent, then, that we ought to be engaged in evaluating each program component at the time that we are engaged in it, as a continuous process. Perhaps we should build in ways of obtaining information from all those involved in the system -- that is, from one another -- about the effects of our own activities?

There is a modern paradigm for working in this way. Many of you, I am sure, have been in human relations labs or workshops in which the whole method of working is essentially setting up continuous feedback systems among participants to be continually evaluating the effects of their own participation. For example, if we are having a committee meeting, then quite clearly we have to be concerned about why we are doing it, what we hope to achieve, and which activities of the meeting are contributing toward the outcome, are moving us toward our objectives, and which are not productive. If that particular meeting is effective, if the next meeting is effective, if each one of our sub-activities is effective, then we have increased the probability of obtaining optimum program outcome.

This raises another issue. If we are effective in all of our activities, in all of the sub-parts of a program, does this imply that we will optimize the outcome of the program itself? Is this not contingent upon whether we have organized these various activities in a way that they interlock in contributing to the program goals? Given that our activities are meaningful, and that they are related to one another in a way that should make for optimal program success, then it would appear that the degree by which we maximize the effectiveness of each particular sub-activity should reflect the degree of program effectiveness.

This is one approach to avoiding the pitfalls and difficulties of trying to evaluate a highly complex program, complex in terms of people, activities, space and time. One way is not to look at outcome, but to build the evaluation into each part of a program, as a continuous feedback process.

Now, in the final analysis, this is what we believe in, isn't it? It would appear that when we are talking about mental illness and mental health, we are talking about communication phenomena among people. We are discussing their ability to act effectively vis a vis one another, to be able to evaluate the effectiveness of their acts on other persons by receiving feedback from them, to be able to perceive feedback and interpret it accurately, and therefore to be able to modify actions in terms of its effects upon others. If that same type of philosophy were applied to the various activities in our programs, the continual awareness of the impact of our activities upon one another, evaluation, and modification of those activities, as a continuous ongoing process, then it might well be that more formal, extensive, scientific evaluation programs would be thought much less necessary and relatively less useful.

THE PROBLEM FOR GROUPS 1 AND 2

A Comprehensive Mental Health Plan for Ames County

Situation

The county government of Ames County is considering applying for funds to develop a community mental health center which will be integrated with existing community mental health resources. They will present these plans to the state mental health authority who in turn will review them with the State Mental Health Advisory Board. It is the policy of the mental health authority that a mental health center should provide services which are responsive to the needs of the community and not just "selected services." Your group has been asked to develop a community mental health plan for Ames County. For purposes of this workshop you may make your focus as broad or as narrow as you wish.

Description of Ames County

Ames County, Oregon, is an area of approximately 3,000 square miles, having a population of approximately 40,000 people. The population is approximately evenly divided between those living in rural areas and those living in urban areas. The principal products and industries of the County are agriculture, lumbering, and manufacturing. The average disposable per capita income for residents of Ames County is in the neighborhood of \$2,000 per year.

Ames County has a number of urban areas but only one of these has a population running as high as 10,000 people. The next largest urban area in the County, 80 miles away, has less than 5,000 residents.

The health resources of Ames County include a health department headed

by a full-time physician. This department is staffed by seven public health nurses, two sanitarians, and two clerks. Among the special programs of the health department is a mental health clinic located in the most urban area which is headed by a social worker with a master's degree. The clinic is providing two major services which are in line with the state law governing the clinics in that they must provide at least two of seven basic services. The clinic provides marriage counseling and child guidance. Other options are treatment of alcoholism, mental retardation, adult psychiatric treatment, mental hospital follow-up, and community inpatient psychiatric treatment.

Ames County has three general hospitals, one of which has 150 beds while the other two have 50 beds each. The County also has an intermediate general hospital with 20 beds.

There are five nursing homes in the County, four of these have between 40 and 50 beds each and one has 10 beds. There are 38 practicing physicians in Ames County, one of whom is a board eligible psychiatrist. The largest urban school district has a single school psychologist.

There is also a 400 bed psychiatric hospital in the largest urban area which treats all types of mental illness. A part of this hospital's program includes an outpatient clinic service which receives patients from 14 county districts. This clinic specializes in marriage counseling and mental retardation.

There is evidence of elevated social disorganization in the county. State suicide rates are 14 in 100,000. Ames County has a rate of 17/100,000. The divorce rate for the state is 3.4/1000 persons. Ames County has a divorce rate of 4.8 per 1000. School dropouts are also elevated. Overall the state had eighteen percent dropout between the 9th and 12th grades.

Ames County has a dropout rate of 21.5%. Legend has it that this high rate of dropout is due to a large Indian population. A reservation is located in the county.

So far as criminal activity is concerned, no good data are available on the amount of crime in comparison to other communities. However, the county has a reputation of being a "rough" community.

By coincidence the population of Ames County has the exact age distribution which is characteristic of the state of Oregon as a whole. The percentage of people receiving public assistance and the principal causes of death within Ames County are also by coincidence exactly those figures which are characteristic of the state of Oregon as a whole.

Report from Group 1 -- James G. Hansen

The task that our resource person, Dr. James Bosch, Anthropologist, Oregon State Hospital, gave our workshop group was that we should develop a comprehensive mental health plan for Ames County. Our workshop members, in looking at the task assigned us and discussing the many ideas evolving out of this assignment, found that we were dealing with three difficult problems:

(a) The community needs and problems were not really known or understood by us as presented in the working paper.

(b) We did not know the community services, their relatedness one to another and to the community.

(c) We did not have a definition or criteria for funding of a comprehensive mental health plan.

It is obvious that with the task assigned us and the problems which evolved out of this assignment the group members encountered considerable amount of anxiety. However, out of the discussion came certain suggestions

for a comprehensive mental health program to be initiated: specific suggestions for service, observations, and questions.

Suggestions for a comprehensive mental health program to be initiated:

1. A vital step is the involvement of the community.
2. A formation of a local ad hoc committee, a county committee including or involving the staff of existing services or facilities (once a program is established this then could be an ongoing advisory committee).
3. Resource persons are needed, for if local people knew how to solve the difficulty they would have done so.
4. Identify the power structure so that it can be involved, or sponsor, or at least sanction the program.
5. Determine the community needs and problems as felt by the community.
6. Identify and understand existing facilities and services, their place in the community, and relatedness one to another.
7. Devise a method whereby the community might be enabled to define its problems and find some clue as to the answer.
8. Designate authority to someone or some group to be accountable or responsible to organize, utilize, and plan the services within the community.
9. Initiate an interagency council, enabling all care-giving persons in the community to get together and help one another.
10. Develop in, at least, a minimal way record keeping for statistics and feedback.

Specific suggestions for service:

1. Attempt to provide continuity of care, if possible, utilizing present services.
2. Provide school social workers.
3. Identify individuals in the more rural areas who might assist, i.e., in suicide prevention.
4. Provide psychiatric out-patient care.
5. Round out, expand present services, possibly utilizing hospital psychiatrists in the community services and clinics on a part-time basis.

Observations:

1. The problems specified in the paper presented to us -- suicide and divorce rate, dropouts, low income, and the low educational level -- might not reflect the real needs or problems in the community.
2. A community defines its own problems, i.e., those who do not fit into its current patterns, or those who cause discomfort in the community.
3. The geography has an effect upon or influences services provided.
4. This county epitomizes program problems in the total state: (a) distances involved, (b) low population density.
5. Experts brought in without local involvement are frequently unsuccessful.
6. Funding programs set up guides, limitations, and controls over programs and services.
7. We cannot understand and interpret these statistics unless we are in the community. We could speculate that this is a more mentally healthy county than others. Should we therefore attempt to keep more marriages intact for better statistics?
8. Although it wasn't presented in the paper given to us, it was found from another source that many persons and services in this particular county are interested in a comprehensive mental health program. Some of those individuals mentioned are the county judge, county health department officials, hospital officials, a physician interested in a problem of alcoholism, and the juvenile officer.

Questions:

1. What is a mental health program? One providing diagnosis and treatment, or education, or a better organization of the services for a county, or all of these?
2. What does this county define or see as its priorities in needs or problems?
3. What is needed? A centralized or decentralized program?
4. Are the needs in this county different from needs in other counties?
5. Why is the county government asking for assistance and no mention was made of professionals?
6. This county has a variety of services. How are they utilized and coordinated?

Report from Group 2 -- William Barber

Our group began with acknowledgement that there appeared to be a recognizable need for Ames County to make some changes. The suicide rate was seen to be higher than for the rest of the state, as was the divorce rate and the rate of school dropouts. It appeared that the population of Ames County tended to account for the high rate of school dropouts, by the fact that there was a large Indian population in the county. However, there seemed to be no evidence to support or reject this idea. We did note, however, that a definition of terms such as "dropout" would be important because what might be termed "dropouts" in Portland would not necessarily constitute a dropout rate on an Indian reservation where scholastic expectations might be somewhat lower. We recognized, too, that there were factors about the suicide rate which we did not yet know. The rate was seen to be a little high for the average of the state but here, too, further investigations seemed called for. I recalled that in some western town there is a Boothill with a tombstone that says, "He called Bill Smith a liar." This man might have been listed as a suicide statistic, under the circumstances.

We felt, therefore, that we needed to know much more about the community we were studying and that there might be some statistics available already and others which might be readily gathered. Dr. Caver suggested, for instance, a narrative study of the population and its history. Demographic studies have probably been done already and there were doubtless numerous resource persons in the county who might be helpful to us. For example, the county agent.

We concluded with six points which might be described as areas for question. First, then, would be a study of population. Where are the people, including the Indians who were mentioned as possibly representing a

separate group? How separate a group are they? How is this particular segment of the population related to school dropouts and suicides? Second, what resources are available now, such as the clergy and the single mental health clinic? How available geographically is it to schools and other members of the population? And who is eligible for service in the clinic? Third, who is using the resources which are available? And who is "getting through the net," as one of our group members suggested? Fourth, by whom will the study be conducted? We first thought of importing an expert but concluded that in theory, at least, we constituted a body called together to begin to explore the problem. So we imagined a sort of lay committee to start with. These are some of the people we thought of who might be on such a committee: a lawyer, a minister, a leading figure from the Indian population, someone representing labor and management, a private physician, a member of the law enforcement team, which might be the sheriff but might also include the judge. We also felt some service organization should be represented and we agreed that membership on this committee should be of both sexes, as we thought of such organizations as the League of Women Voters which can be a real and potent force in some areas. In addition, we suspected that there might be a representative from the press, the Public Welfare Commission, the area's representative in the legislature, and perhaps some others of whom we had not thought. This group, we thought, might do the initial study or, at least, the initial stages of the study.

The fifth step, we then decided, would be to make known to the community the needs which this investigative body concluded were present, and then get some sort of "feedback" from the community as to what they wanted to do about it. We recognized that unless a program had the broad support of community, it was doomed to failure before it began.

Our sixth and concluding point was that if this study should result in "a comprehensive mental health plan for Ames County," an ongoing committee should exist to "evaluate the innovators."

THE PROBLEM FOR GROUPS 3 AND 4

Planning an Evaluation of an After-Care Program

Situation

Mid-State Hospital is an old established 900 bed psychiatric facility located in a semi-rural area of the state. Its patients come from widely distributed areas connected by a modern highway network. The staff at this hospital looks at the hospital as an intensive care facility for short-term patients, but in fact there are a number of patients who have been in residence over five years. It serves all types of patients, running the gambit of mental illness from the neurotic to the retarded. No special screening service is used. It offers an outpatient service which is integrated into the general program of the institution. Physicians may provide both inpatient and outpatient treatment services and could follow a patient both before and after inpatient service.

Though following the national trend toward reduced inpatient load, admissions have been at a rather constant pace of approximately 800 per year; 650 first and readmissions, and 150 returns from trial visit. Releases have been at a somewhat faster pace than admissions with approximately 950 leaving the institution for community living.

The staff at this institution includes two board certified psychiatrists and one board eligible psychiatrist. In addition, there are ten physicians covering a range of medical specialties. Four clinical psychologists are in attendance, none of whom have research responsibility and three social workers. Nurses, psychiatric aides, and ancillary service workers make up the remainder of the 400 employees.

Community resources are minimal. One community clinic is located in the closest urban area approximately 25 miles away. Its service is primarily child guidance. Public health nurses are available but because of the distances involved, their services are limited.

From June 1963 to June 1965 the resident population dropped from 1,250 to 850. Most of this reduction has been attributed to a shorter length of stay brought about by early release. The hospital staff feels there is empirical evidence to suggest that involvement of the treatment staff with the families of patients and the community in general, prior to and after the release of a patient allows early release and improved adjustment of the patient to community living and reduced readmission rates to the hospital. As a result of the empirical evidence, the hospital would like to set up a study to determine the effect of an enriched program which would provide for pre-release family and community planning, and post-release care.

The plan envisions the addition of personnel with social worker skills to the hospital staff. They will not participate directly in the treatment process in the hospital. They will be available to interview patients and families on admission, at discharge, and after release. They would also be available to visit families or patients in the community. The plan allows a research and evaluation component of sizeable proportions.

* * * * *

Report from Group 3 -- David G. Berger

A problem that we tangled with in our group was the problem of criteria. The protocol provided for our group listed three criteria, i.e., ways by which consequences of the program could be assessed. As we began to explore these criteria it became obvious that there were doubts about the meaningfulness of these criteria in the members of our group. That is, there were

doubts about the relevance of a criterion such as the length of stay as a meaningful measure of assessing the program in the institution. Similarly, there were doubts about the return rate as means for assessing the program at the institution. In spite of those doubts we had to recognize that, if we were going to be in the business of evaluating, we would have to come up with some criteria, whether they would be the criteria that were proposed in the protocol circulated to our group or some that we ourselves were able to develop. I think we also recognized that there is a certain amount of fallibility or difficulty with any criteria that would be generated to evaluate this program. Obviously, one of the real skills or arts in doing evaluative research is to somehow manufacture criteria which are the most useful for getting the kinds of answers that you would like to have.

We also noted that the process of evaluation is in many ways a process of innovation. That is to say, when you really sit down to evaluate, there is a great need for innovating or, putting it another way, when you begin to understand what it is you want to accomplish you find that there are a multiplicity of ways to measure the extent to which you have accomplished it. It is in this sense that our group recognized the kind of innovational skills which are necessary in program evaluation. And I think we identified a number of problems which make it difficult to innovate in the process of evaluation. For example, one of the problems that we came across was the distinction that is too often made in talking about the hospital and the community, as if in fact they are two different kinds of animals, and that they are not somehow intimately related to one another. We raised questions for ourselves, time and time again, as to whether it is legitimate for persons who are supposedly working in hospitals to do certain kinds of work in the community. There were then a number of roadblocks to our thinking as we

went along simply because we all bring with us certain kinds of notions about how far the hospital extends or how far the community extends and whether or not you can somehow breach these boundaries.

Another kind of inhibition to innovation with which we dealt is the problem of whether persons who are involved in treatment can also be involved in data collection, and where the boundaries are for these two functions. We thought that persons involved in these functions could more reasonably be placed on a continuum from treaters to data collectors, and that it might be artificial to talk about staff who have one function exclusively.

Another kind of problem we talked about is the whole issue of state laws and policies and the extent those state laws may limit innovation in research and evaluation designs. For example, we had some problem with the question of whether within the state of Oregon at the present time it is possible for persons who are working in institutions to get involved in working with the patient's family. There were a number of reservations about this. Yet a number of the members of our group felt that this might be an essential component of the treatment modality that we were attempting to define. So that we felt that recognition must be given to state laws and the role they could play in limiting the development of programs and the criteria for evaluating the programs.

I might take a few minutes to show you the kind of model that we developed. We have it on the blackboard. It's kind of messy, but in essence what we were talking about involved two problems in the project that we were handling. We were talking about the problem of how to develop criteria or goals by which to evaluate the program, and we were concerned with the problem of how to define the program we were asked to install in the

institution. We spent a great deal of time enumerating the approaches for both of those components of our evaluation design. In looking at this now we can see that it's really impossible to talk about a simple problem in evaluation. There are apparently no simple problems. Any evaluation problem that you look at is a very difficult, intricate problem, if you really try to define what's involved in understanding a program. For example, we spent a number of hours discussing the various components of what would be involved in the work of additional social workers, how they would actually work in the community, how they would work in the hospital, what kinds of relationships they would have with the family, with the patient, with the community worker, and with other staff working in the hospital. We spent a great deal of time talking about what the notion of release means in understanding hospital performance. We spent considerable time talking about the return rate and what it would mean, as well as what adjustment to the community might mean. It was clear to each member of the group that each of these possible criteria has significant complications attached to it, that each needs a great deal of thought before one would feel comfortable in using them in any research scheme. Essentially I think it was quite clear to us that unless one took the trouble to really involve the whole institution in a project like this that the project could fall on its face very easily.

In summing up our experience I do want to say that I thought the group really had a good time talking about this. There were a number of suggestions that came out of the group discussion with respect to the kind of experience we had. One of those suggestions which I think may be of real value to you is that this kind of exercise that we went through might be a very worthwhile exercise for the staff in given institutions to participate

in with some regularity. That is, if we as a group coming from a number of institutions could really deal with the problem like this and get a great deal out of it in terms of understanding our own institution and our own roles in our institutions, that it should be extremely useful for each of our institutions to somehow feel the need not only to ask individual staff to think about their roles in the process of evaluation, but to encourage staff to use staff meetings, from time to time, to think about the institution's role in the process of evaluation.

Report from Group 4 -- Wesley B. Terwilliger

I think our group would concur with the findings of Group 3 as reported by Dr. Berger. We approached the problem a little bit differently, however. Our group initially was a little more bound by the plans that were presented to us. We had some difficulty with that plan. We had difficulty getting started because we were not quite clear about what was being said. We weren't clear what the intentions of the plan were exactly and we did note a number of contradictions in it. We weren't sure whether some of the goals were primarily hospital oriented and some of them community oriented or whether they were both. We had to discuss this at some length and to some extent leave the original plan and begin our thinking all over again. We reached the same kind of conclusion that Group 3 did, that is, that it's a very complex piece of work and research. As we looked more into some of the meanings of what we were trying to get at we found that there were so many facets involved in each approach that it would take between six months to a year to really design a good, thorough, meaningful kind of program.

We, too, picked up the criteria that were listed on page one of this program and we chanted them a little bit. We talked about what the nature of after-care was and what its goals were. We tried to see if there were a

distinction between after-care as it related to the hospital and after-care as it related to the community or whether these two were essentially the same kind of thing. The general plan that we came up with was a very vague kind of thing and it runs something like this: Increased use of therapeutic staff in the community. When we talked about the goals as they are listed on the sheet, the first one was to affect early release. We decided to change that to "facilitate optimal release based on patient need and a knowledge of community resources available for him." The second goal given was "improved adjustment of the patient in the community," and we accepted as a goal "improve the adjustment of the patient to the community and of the community to the patient." We felt that it had to be a two-way kind of thing. Then, for "reduced admission rates to the hospital," we again used the term optimal and said "optimal readmissions to the hospital," meaning by that, that when a patient came back it would be an appropriate readmission to the hospital and we weren't so much concerned with the rate of returns as such.

We talked quite a bit about what was meant by after-care and decided that it would vary in a number of different settings. We felt that a number of variables were affecting it and that some of these variables were characteristics of hospital populations which in itself would differ between different geographic and cultural areas. We felt that social variables in the community had a great deal to do with the kind of care that we would provide. When the statement on the sheet said that there were minimal services available in the community, we didn't know quite what that meant. It did list a public health department and that's about all, but we wanted to look more closely into what really was available to the returned patient or to the pre-admitted patient in the community. We felt that there would

be a need to investigate all the social agencies and facilities available in the community, the vocational climate in the community to which he was returning, what kind of supervised living facilities were available, were there foster home, half-way houses, day and night centers, what kinds of outpatient care were available for the patient. Were emergency services available, what kind of medication control would be available, were there any kinds of child guidance, any kinds of therapeutic relationships or consultation available to them? Were there welfare agencies, what was the attitude of the police for the hospital patients, what was the character of the schools in the area? Were there vocational people? Special education people? What about the physicians in the area? How did they feel about seeing hospital or mentally ill patients?

So the conclusions that we came to yesterday after our first group discussion was that after-care programs depend upon the nature of the hospital population and the tolerance of the hospital and communities for particular types of people. We recognized that some kinds of people were released earlier than other kinds of people. What was available in the community, what was the nature of the patient's family? What were the administrative pressures on the hospital to release people? Where would the money come from to provide this kind of after-care service? Would the resource from which the money came affect in any way the kind of program that we could conduct?

Today we took up the general topic of evaluation. Again we restated the goals that we modified from our worksheet. The general agreement was that a number of things would have to happen in order to evaluate a post-hospital or after-care service. First of all, we agreed that we would need a large sample of patients in order to come to any solution. We would have

to look at these patients in two different ways, both subjectively and objectively. In our subjective appraisal of the patient, we said that we would want to talk to the patient himself after he had been in the community for a period of time and ask him how he felt he was getting along. We would want to talk with his family and see what they felt about the patient. We thought that we would go to his employer and see what his employer felt about his adjustment. We would want to talk to his neighborhood, his relatives and his friends and any other kind of variable or factor we could think of that would help us get a subjective opinion of the adjustment of this patient to community living.

Objectively, we decided we wanted to look at a number of other kinds of things, some of which were fairly readily available. First, we would want to look at marital status, how did this person adjust to his marriage. Did he divorce? Was he separated? We would want to look at his employment record and see how many times he changed jobs, how long he held the jobs, and what his level of skill or work was on these jobs. We would want to go to the courts and see what kind of contact these patients had with the courts. How many suicides occurred among the patients? How many murders? How many burglaries? How many of these kinds of involvement? We would also want to go to other agencies within the community and see what the nature of the patient's involvement with them was, how able he was to make use of these agencies. We would want to go to his physician and get some kind of estimate of his physical well being. Then we would want to give some kind of objective test to the patient and perhaps others to get some sort of indication on a statistical level of what sort of adjustment he had made. The point was brought out emphatically that we would need these data in two different areas. First, we would want this kind of measure on the pre-morbid

individual, that is, what had happened to the patient prior to the time he came to the hospital. And we would want to get it then after a period of time following release. We weren't too specific about how long but we did say we would want to get pre- and post-evaluation of this patient.

In talking about design we talked about the need of gathering a great deal of epidemiological data. We would want to analyze what had happened in the hospital prior to the inception of the follow-up program. We would want to look at what had been the length of stay in the hospital, what had been the length of stay of released patients in the community, what the re-admission rate had been, what the route was by which patients had been coming to the hospital, that is, voluntarily, court committed, emergencies, and so on.

We talked about a number of decision-making processes that went on in both the entry route and the discharge route of patients. We wanted to find out the variables affecting the decision to release the patient from the hospital and what kinds of things affected the decision to come to the hospital or to be committed to the hospital. We wanted to have some evaluation of the tolerance level for mentally ill patients of people both in a hospital setting and in a community setting.

In summary, we decided that we wanted to examine a number of variables both in the community and the hospital. We felt it desirable to have measurements on the patients pre-intake adjustment and post-discharge adjustment. These measurements would be obtained both subjectively and objectively.

THE PROBLEM FOR GROUPS 5 AND 6

An Evaluation of a Program for Defective Delinquents

Situation

A state institution for the mentally retarded has responsibility for the care of a large number of mentally retarded whose aggressive behavior and emotional disturbance is such that they are classified as defective delinquents. Recently a near riot has led to the development of a new program for treatment on one ward (Ward A). That treatment program has a basic and an expanded program. After considerable experience, the program is to be expanded to a second ward (Ward B). The goal is to demonstrate that defective delinquents will respond with more mature, responsible behavior to a treatment program geared to their level of functioning.

Facilities Available

Each ward has approximately 11,000 square feet of living space. There are 24 beds on Ward A and 60 on Ward B. There is a day room on each ward. On Ward A part of the sleeping area has been converted to a dining room. This area was set aside for this purpose in 1960 to avoid the "perils" of moving the more aggressive male patients to the regular dining area in the basement of the cottage. A fenced outdoor recreational area is available off Ward A. Toilet, bathing, and clothing storage facilities are present on each ward, as well as a barber shop and a nurses station. There is also an area on Ward A set aside for a craft's workshop, academic instruction and family meetings.

If the project is approved, an area presently used for storage on Ward B, would be converted by the staff to an area suitable for a ward craft

program, for academic instruction and for family group meetings. One thousand two hundred square feet of space on Ward A will be remodeled and equipped as a woodshop.

Staffing

The ward staff will be enriched by the addition of social workers, aides, a vocational instructor, a recreational therapist and a psychology consultant. A teacher is available full time. A psychiatric consultant is available.

Patient Population

During the period of the demonstration program we expect to carry 24 patients on Ward A. The age range on this ward is 17-35 with an average age of 24. The I.Q. ranges from 40 to 74 with an average of 58. Patients on Ward B will be comparable for age and I.Q. They will number 50-55 and will exhibit behavior problems of a somewhat less intense variety. All patients on the two wards will nevertheless fit the definition of defective delinquent as employed for this program: those borderline and retarded male patients who are so dangerous to themselves, to the community, to other residents or the staff, that they must be treated as security risks.

The Basic Program

The ingredients of the basic program are: (1) to establish a common understanding about ward operations through an ongoing program of ward meetings, involving all staff in one series of meetings, and staff together with patients in another; (2) to assess patient needs by (a) a program of personality evaluation through psychological testing and medical appraisal and (b) by establishing social service contact with family or friends in the community; (3) to provide individual counseling to a limited number of patients who have prospects for profiting from such help; (4) to provide a

limited craft-educational program designed primarily to keep patients occupied and provide a beginning sense of achievement.

In the first steps of the therapeutic approach, the staff and patients faced each other with their mistrust. The sessions sometimes ended in shouting and threats of suicide. Sometimes patients needed to be restrained after the sessions because of their aggressions. But the staff patiently pointed out that it was this type of behavior which prevented the staff from granting the privileges the patients desired. The therapist kept repeating that they could control their behavior and that when they did, they would be rewarded. The sessions were unstructured in that no plan was followed but each question and problem was taken as it came.

The staff granted requests only when all felt comfortable with the change. The staff agreed together on the limits which would be imposed upon patient behavior and made these limits explicitly clear to the patients. The increased communication between day and evening attendants and the exposed treatment methods initially created some anxiety but soon this same openness became important for confronting the patient with his behavior. As patients began responding, staff began rewarding the desirable behavior. Sometimes the whole ward was rewarded; other times individuals were given rewards.

To further implement the positive relationships which began to develop, individual therapy began for a few patients. At first the psychologist met with one patient and the charge attendant. As identification took place further, the psychologist felt comfortable in having the attendant meet alone with the patient.

The staff focused and responded to those behavioral aspects displayed on the ward which have made the individual unacceptable in other living

situations. In responding to this behavior they have:

1. Given instructions that are explicitly clear at the patient's level.
2. Explained carefully why restrictions exist.
3. Explained how restrictions may be removed.
4. Made goals short term and rewards tangible and frequent.
5. Pointed out mistakes positively by showing the patient he has other positive aspects that we accept.
6. Used as much repetition as needed.
7. Directed patients energy into meaningful work experience.
8. Increased individual tolerance and responsibility concomitant with the improvement demonstrated.

The Expanded Program

Under the expanded program, patients will be organized by age and intellectual ability into treatment groups of about 5 or 6. They will then be involved, insofar as possible, in various aspects of the intensified treatment activity as members of these small groups. Each aide on the ward will have a small group assigned to him for intensive contacts throughout the day. This will enable the staff to adjust the demands of the program more directly to patients' needs. It is planned to establish a schedule for small group therapy where all patients will be involved in their small treatment groups. These group therapy experiences will be under the direction of the physician, a social worker, the teacher, or the psychologist. Each group will also have an aide acting as an observer-participant in the beginning and later as a leader-participant. Aides will be rotated through this experience.

The academic program will be expanded to offer class activities one hour per day organized again for small treatment groups. Coupled with this

program will be an expansion in the crafts activity emphasizing the assembling and construction of simple useful objects -- such as toothbrush and clothing racks. This aspect of the ward program will be accented for groups containing the slower, less capable patients, while the more intelligent patients, who need specialized help, will be exposed to heavier academic schedules. The small treatment groups will also be scheduled as units in a manual occupational training program. Under this program the group will be supervised by the vocational instructor who will work under the guidance of the project teacher. The additional aide for Ward A will assist the vocational instructor in the expanded occupational training program. The purpose of the program will be to imbue patients (1) with realistic attitudes toward occupational supervision, (2) with greater awareness of the needs of other patients working with them, and (3) with sound regular working habits. Work assignments will involve outdoor maintenance and repair activities as well as indoor woodshop activity. The latter activity will also be a constructive program with definite work objectives since it will serve as a repair shop for institutional furniture. The general format will be to grade work assignments to the level of the patient and to make certain that the patient gets some feeling of accomplishment for each day's work.

There will also be other individual and total group activities for the patients. Patients who can profit from a more intensive relationship will be seen in weekly individual counseling sessions. All others will be seen individually each week to review their progress. On week-ends the recreational therapist will instruct aides in leisure time recreational and athletic activities.

Two other components will be introduced with the expansion. They will involve a patient-placement program and a family counseling program. The

social work staff will be largely responsible for these services. They will be assisted by the vocational counselor in the matter of placement. Families will be invited to visit the wards more regularly. They will also be encouraged to participate in bi-monthly group sessions conducted by social workers for patients' families during week-ends. It is planned to attempt to bring patients into some family sessions. Hopefully these two aspects of the expanded program will help the ward staff appreciate the realities of community attitudes toward the patients, as well as provide community residents with a more objective understanding of patients' potential for productive instruction and community life.

* * * * *

Report from Group 5 -- Joseph Treleaven

Our responsibility was to review a project at an institution for the mentally retarded for defective delinquents and to come to some conclusion as to how this project might be evaluated. I will give you a running picture of how the discussion went.

At first most of us felt the need for more information about the project and we asked a member of our group who works at the institution innumerable questions about the project. This project was developed after a riot in a cottage at the institution back in 1962. It appeared that the riot was the result of a previous project. The previous project was designed to transfer all troublesome characters from the main campus at the institution to a place known as Prig Cottage several miles distant, with the hope that once out of sight they would be out of mind. Apparently this did not work, because after a certain amount of time the staff had to be more and more restrictive and the patients became more and more deprived and unhappy with the circumstances of their lives and finally rioted. Thereupon the

institution and the mental health division responded by pouring in considerable resources. They added personnel of various sorts, they developed programs, and finally it came to a grant from the federal government to carry on this work and expand it. So we asked ourselves why do you want to evaluate this program? After a lot of discussion, the representative from the institution said that they had to evaluate it, that it was part of the terms of the grant. So we established one of the reasons. However, we also established that we wanted to improve upon this project. Even though things seemed to be coming along very well now, there is still the hope and expectation that there might be ways to improve it. There might be ways to more efficiently and effectively use the resources we are pouring into this project.

Then we raised the question: Evaluate what? Or how could we evaluate it? What were the goals that we wanted to evaluate as to whether or not we had attained them? As we discussed this, it became apparent to us that there were two different goals and objectives that were somewhat different. The first is promotion or mere justification. We have something here that we think is good, that seems to be doing well, and we would like to convince others of this and obtain continuous support for this type of activity because we think it's a good thing. However, we have another very essential objective. We came to the conclusion that we should evaluate this to determine what it was that we were doing in this project that was essential to the success of the project, what it was that produced the desired change in the patients involved, and whether or not this was to their benefit.

If you have done certain things in a project and have a certain rationale -- the rationale primarily being that by improving the communication among the staff and between the staff and the patients and by improving

the skills with which the staff handled disturbed behavior or delinquent behavior -- the amount of behavior disturbance in these patients would decrease. However, we also considered another hypothesis which might be equally valid: Because of the pouring forth of interest, money, personnel, and involvement of everybody at the institution, of the mental health division, of the board of control, and to a certain extent the public at large, there might be a non-specific activity that produced the given results. So then we asked ourselves this question: How do you design an evaluation project to separate out non-specific effect from that produced by the specific training of the aides in communication?

From there we progressed to the idea that the best thing to do was to either measure change in staff attitude or to measure change in behavior. I think the discussion went more into measuring change in behavior and we discussed for a considerable length of time how this might be done. One of the concrete suggestions that was brought out and very well supported was that we needed to take samplings of concrete behavior, record them according to a check list, and indicate either that the behavior was there or it wasn't. This way we might develop some means of evaluating behavior change. This ended the discussion for the first day.

The discussion started on the second day with the suggestion that evaluation was a communicative process and that evaluation consisted of passing on information. We had to consider in so doing what type of information we wanted to collect and pass on, who we were to pass it on to, and how it was to be used. It is necessary in any evaluation to consider the extent to which we will go to collect and disseminate information. We raised the question as to whether or not evaluation should be carried on concomitant with program development. Could we have a constant feedback which would be

used to adjust the program? It was pointed out that this was informally done in most programs. We raised the question as to whether or not this should be more formally done in a way which would help us eliminate some of the biases and tendencies that are all too natural as we try to justify our efforts rather than to see the truth in what we are doing.

In discussing this I reported that it was brought to my attention that some of the cases that had graduated from this program at the institution for the mentally retarded had subsequently been admitted to the state hospital, and if one were to follow the course of this person's life history over an extended period of time, beyond release from the institution, one might raise very different questions of the value of this program at the institution then if one merely assessed him at the time he left contact with the institution. This points out the real problem of evaluation: to extend it beyond the scope of the institution into the life span of the individual.

Then we asked ourselves: Could we evaluate this project at the institution by comparing the results there with a matched control group from another institution, say a group of delinquents or defectives in other institutions? Someone suggested that the WICHE data processing work that had been done in recent years might be of assistance in this.

Report from Group 6 -- Fred E. Letz

The group that I was in, first of all, came up with many of the things Dr. Treleaven said and we, too, decided that sometimes it is dangerous to work without sufficient knowledge. In this instance, perhaps we did not have enough. That reminded us of the little violin player walking through the forest who was confronted by a hungry lion. He was frightened and he did not know what to do but to start playing the violin.

The lion heard the soothing music, stopped the attack and fell in behind the little violin player. They wandered on through the forest. Pretty soon they came to a panther, and the same thing happened. The panther gathered himself to spring, the violin player played beautiful music, and the panther was soothed and fell in behind the lion. The three of them continued on through the forest. After a bit they met a tiger. There again the tiger gathered himself to spring, the violin player started to play the music and the tiger sprang on him, killed him and started to devour him. The lion said to the tiger, "Why would you do this horrible thing? Why would you kill something that has the ability to bring this beauty to the forest?" The tiger, placing his hand behind his ear said, "Eh?"

The first thing we did in this group was to decide that in the charge given us on this paper, the key word was treatment, that is, to see the effect of treatment upon these delinquent patients. We also decided that as one looks at this, it really could be divided into two separate programs. Program I in Ward A was already in existence and that to evaluate it in the way that one might like to evaluate would be an impossibility. One could only evaluate in the sense of the total program. One could say whether there was a change in the behavior of the patients or whether there was not. One could draw up certain criteria to see what thing change might be. We made a list of such things as, after the program started, how many of the patients reported into the hospital banged up from each other. One might check on their behavior as to how much furniture was broken, how much hostility was shown by destruction of other property, by the expenditure of paper towels in the lavatory, and other such indices. One also might try to check as to what kind of behavior changes there were in the more positive aspects of whether or not there was more communication between them, whether

or not they entered more easily into games and had better relationships. You could say then, in effect, that obviously something in this program seems to have had a positive effect. We could not say whether the effect was simply a matter of their getting more attention, or whether it was the effect of different programs that were brought into it. It was agreed that the greatest significance would probably be the change in the attitudes of the aides and the other people working with the patient. Here again this might not be something that could be well measured at this point in the program. This being disposed of we decided to think of Ward B as a separate project. In this instance an operational definition of delinquency, methods of screening the aides, methods of evaluating aides and participants and patients would be needed. There are schedules that exist or others could be drawn up to make this evaluation. We could keep some kind of a continuing control or continuing check on *attitudinal change in the staff and in the patients*. In order then for there to be any really discriminatory decisions on what was effective, we agreed that you had to set up different groups. There had to be some kind of a control. We thought given this number of people, sixty, with the stated age range, the group would lend itself very well to about twelve different groups with different I.Q. levels and different ages.

We assumed which was our prerogative since we had no information to the contrary that the model distribution of ages would be just right. We would come out with five people in each group. From this basis, it would be possible to set up control groups, either keeping them in the same setting but with only half of them getting treatment, and by treatment, I mean the total thing of counseling, group counseling and occupational therapy as opposed to the regular institutional program.

Here again we felt that under either circumstance the evaluation of the staff would probably be the most significant factor that we could come up with. There certainly was no disagreement among us that there would be change, because anything one does differently with these kids is going to bring about some change. We would need to keep track mostly of change in staff attitude.

One of the things that came up toward the end of the discussion that I thought was fairly significant was that one could certainly develop schedules to show change, in terms of behavior in both the aides and the patients. Another was that one could so model a program to bring about the desired change as well as undesired change. There were three statisticians in the group. One of the things they said which I think was very good was that in their experiences statisticians are called in after one has set up a program and after it is decided what information to be shown. One then comes to them and says, now devise me some means of showing this. They were saying that if they are to be used really well, they should be in on the planning from the beginning. You gather the data and these data will show something. It may not necessarily show what you want to be showing, but at least it will probably be more valid information.

IMPLICATIONS OF THE WORKSHOP

E. Myles Cooper
Program Analyst, San Francisco Regional Office
U. S. Public Health Service

My function is to discuss the implications of the workshop, or what the logical inferences are that each of us may draw from what has gone on here the last two days. Each of us will find it necessary to relate these discussions to our own particular situation, whether it be in a hospital setting or in a community setting, whether our perspective be clinical or research-minded, or to the variations in our disciplines.

A number of people have mentioned the purposes of evaluation. Dr. Neal pointed out that it really gets down to who needs what, when and why. Then we added a couple of adverbs of how it's to be done and where it's to be done. Dr. Treleaven mentioned this afternoon that the purposes may be political or practical, or they may be philosophical as Mr. Jasperson mentioned yesterday in a search for truth. So we have a large spectrum of purposes.

I don't think a detailed definition of evaluation is essential here, since I think we are all quite clear that evaluation means any assessment regardless of methodology or regardless of orientation. It was particularly appropriate yesterday that Dr. Neal spoke from a management perspective and Dr. Saslow and Dr. Jackson spoke from a clinical perspective. All methods are included when we speak of evaluation whether they be experimental, using comparative methods, or whether it be a simple search for an explanation of a phenomenon. Of course, in the latter circumstance, it's frequently inadequate because our judgment as to the reasons for the phenomenon may

not be satisfactory. For this reason, there is great reliance on the scientific method, the experimental method. The principles of evaluation are clearly applicable to both the institutional and non-institutional study. This was implied by the difference in the nature of the problems that were presented to us in the workshop groups, although not specifically alluded to.

During the discussions there were a number of references made to what some of us consider to be sequential steps in evaluation. I will enumerate them and please feel free to expand or criticize them. As I indicated in the beginning, these are the inferences which I drew from my perspective and the inferences which each of you drew may be entirely different.

Step No. 1. Define the situation. This was referred to over and over again. Unless we define the community, the environment, the program, it is extremely difficult to carry out an evaluation program. Dr. Berger mentioned particularly the importance of understanding the environment.

Steps No. 2 and No. 3. Define the goals. We might think there are two sets of goals -- philosophical goals and operational goals. It is difficult for me to conceive carrying out a sound evaluative program unless there is a clear understanding of both the philosophical and operational goals. Dr. Neal and Dr. Jackson pointed out yesterday that unless the goals are defined, there is no basis for evaluation, and this afternoon this was mentioned again by Mr. Barber and Mr. Terwilliger. We all clearly recognize that this is one of the first major steps in evaluation -- identification of the goals.

The goals may be either long range or short range, obviously. Dr. Saslow pointed this out, and again we need clearly to recognize what part of a program we are attempting to evaluate and what the particular objectives are in this respect. There may be other non-program goals as was

mentioned this afternoon, and it's important for those of us who are involved in evaluation to recognize the goals regardless of what they be -- political, philosophical, personal, practical -- and the extent to which they have an influence on the program or on the evaluation.

Step No. 4. Questions. Yesterday Dr. Neal and Dr. Saslow mentioned particularly the importance of defining the questions to be answered. If we are to select data which are pertinent to the program and to the identified goals, the particular questions must be sufficiently clear so that the designers of the evaluation program and the people involved in carrying it out have no problems in communication.

Step No. 5. Identify the protocol. This is the central part of evaluation. The projects assigned to us in the workshops intended that protocol be identified, and we considered the factors which go into such identification. We began to recognize the number of problems involved, whether they were problems in conceptualization, environment, or resources. The number of variables which might be effective in a given situation was noted just a little while ago. The controls were mentioned both yesterday and today. The availability of money, the availability of time, the adequacy of measuring instruments, are all factors which were mentioned during the past two days, and they all relate to the establishment of the protocol for the evaluation of the program. In this respect, if I may be redundant, I want to refer to the importance of definition because it was discussed in so many instances.

Dr. Neal and Dr. Jackson also pointed out that data need to be useful, not just interesting. However, Mr. Jasperson noted that sometimes when we search for truth we do not always know exactly the use to put the particular data and therefore occasionally it may be necessary to include some data

in the design where we do not have a clear idea of how they may be used. This harkens back to the problem of conceptualization.

I can think of one particular instance offhand where the collection of data peripheral to the conceptualization or to the hypothesis was important. Dr. Martin Wolins of the University of California some years ago conducted a study to determine the basis for foster home findings in child welfare agencies. The hypotheses were developed by workers in some five agencies. Having a computer available, Dr. Wolins built into the design the collection of certain other data which he thought might be relevant. As it turned out, the actions in foster home findings in five particular agencies were found to be more correlated with some personal attributes of the social workers than with regard to policies of the agencies in foster home selection. This exception, in my judgment, indicates that the rule that our object should be utility of data and not interest of data, although generally valid, needs not be adhered to at all times.

A. Alternatives. In at least one group, the importance of identifying alternative protocols was discussed. Only a consideration of various alternatives makes it possible to select the optimum. This is due to such factors as: some protocols may be too costly, or take too long, or adversely affect the patients, the staff, or the community, or be otherwise impractical although scientifically appropriate to evaluate a given procedure. Thus the protocol selected as being optimum in the circumstances may represent a compromise between theoretical efficiency and practicality.

B. Prospective. The need to develop our evaluation plans in advance of data collection has been mentioned several times. There are many instances when it is impossible to go back and collect the data to appropriately evaluate the program and this is one reason research grant programs

require built-in evaluation. This is particularly true in a before-and-after situation. Mr. Letz mentioned this afternoon the very common problem of bringing in a statistician to analyze data which have been collected without his concurrence and which may not be germane to the objective of the evaluation.

C. Criteria are judgments. I was particularly interested to note that Dr. Berger, Dr. Jackson, Dr. Saslow, all of them, made a point of the importance of judgment. The identification of a criterion is a judgment. There is no getting away from it. Dr. Saslow discussed yesterday a number of treatment methods discredited due largely to the fact that the judgments at that time were based upon inadequate knowledge. Let me illustrate.

The selection of a base date is a simple judgment which may be important in an evaluation. The definition of improvement sometimes may be completely objective, based upon hard criteria, but nevertheless the selection is a judgment. Frequently these judgments may be very arbitrary, but there may be many situations in which an arbitrary decision is the only one which can be made. This is an area where each of us has equally good judgment, but the evaluator or the person making the ultimate decision obviously is the one whose judgment needs to prevail. On a retrospective basis it may turn out to be bad, but this is one of the real problems and this is the reason so many of the speakers emphasized it.

D. Constraints. The various constraints which are applicable to a given situation were mentioned over and over again, whether they be the social structure in the community, the amount of money or time available, the staff available, the orientation of the evaluator to the program, the adequacy of measuring instruments, the patient's desires. All of these are constraints. As a result of these, it is necessary in many instances to

set priorities in the methodology or in the allotment of resources, as Dr. Neal implied yesterday. Dr. Jackson also pointed out that we must recognize and consider these practical constraints, but that we want to maintain objectivity as far as possible so that there is a minimal effect upon the evaluation. Of course, some of the constraints are much more subtle than those we mentioned. In any particular situation, the evaluator in his conceptualization or hypothesis needs to give consideration to as many of the constraints as possible.

Step No. 6. Operate. There wasn't much discussion of the point of operation of the particular evaluation procedure, whether it be experimental or otherwise. I think it inappropriate to spend any significant amount of time on it at this stage.

Step No. 7. Analysis and Recommendation. Some of the factors of analysis, aside from the constraints mentioned earlier, were noted, whether they be the enthusiasm or dedication of the innovators or evaluators, the personality of the people involved, their personal commitment to the program, and other similar things. As Dr. Neal implied yesterday, in the analysis of a process it is necessary to look both at its effectiveness and its efficiency. Effectiveness refers to the ability of the procedure to approach the identified goals. Efficiency refers to the utilization made of the resources which are available.

Step No. 8. Act Upon and Adjust Both the Program and the Evaluation. This was touched on briefly this afternoon by Dr. Treleaven. The question was raised as to whether adjustments should be concomitant with the on-going program and whether they should be formal or informal. Obviously there are different opinions in this area. The point should be recognized, of course, that unless changes are on a formal basis, it becomes more difficult to

continue the evaluation which is either a continuing process or a sequential process. However, unless program changes are made formally so you can recognize the change, evaluation becomes extremely difficult.

Communications were discussed throughout the workshop, and, of course, communications are significant throughout the evaluation process, whether it has to do with the conceptualization, the objectives, the understanding of the instructions, or the feedback of the results. Obviously, clarity and understanding are especially important in the communication process.

There is a key word in evaluation, I think: Why. It seems to me that "why" is the ultimate philosophical objective in the evaluation process. Care must be taken to understand the real "why" of a situation before making a decision. Isn't this what we are after in evaluation, to be able to understand the causal factors in any given situation so we can draw appropriate conclusions and take appropriate actions?

ROSTER OF PARTICIPANTS

Gaynelle Alfred, Director
Social Services Department
Dammach State Hospital
Post Office Box 38
Wilsonville, Oregon 97070

Frank Bales
Assistant Secretary
Oregon Board of Control
Salem, Oregon 97310

William Barber
Psychiatric Social Worker
Multnomah County Division of
Mental Health
12240 Northeast Glisan
Portland, Oregon

W. W. Benson
State Registrar
Idaho Department of Health
Boise, Idaho 83701

Dr. David G. Berger
Institution Research Administrator
Oregon Board of Control
Salem, Oregon 97310

Dr. James W. Bosch
Research Office
Oregon State Hospital
Salem, Oregon 97310

Dr. Willis H. Bower
Deputy Director of Research and
Training
California Department of Mental
Hygiene
Sacramento, California 95814

Dr. Dean K. Brooks
Superintendent
Oregon State Hospital
Salem, Oregon 97310

Joan Calef
Medical Records Librarian
Dammach State Hospital
Post Office Box 38
Wilsonville, Oregon 97070

Dr. D. L. Callicrate
Clinical Director
Fairview Hospital and Training
Center
2250 Strong Road, S.E.
Salem, Oregon 97310

Dr. John D. Cambareri, Chief
Program Development and Research
Idaho Department of Health
Boise, Idaho 83701

Dorothy E. Collard
Mental Health Nurse Consultant
Mental Health Division
5 State Capitol
Salem, Oregon 97310

Calvin C. Cooper
Chief, Biometrics Section
Oregon Board of Control
Salem, Oregon 97310

E. Myles Cooper
Program Analyst
U.S.P.H.S., Region IX
Federal Office Building
San Francisco, California 94102

David H. Ellis, Director
Jackson County Family and Child
Guidance Clinic
1313 Maple Grove Drive
Medford, Oregon 97501

Harold E. Free
Chief Psychiatric Social Worker
Tualatin Valley Guidance Clinic
13777 S.W. Farmington Road
Beaverton, Oregon 97005

Dr. Kenneth D. Gaver
Administrator
Mental Health Division
5 State Capitol
Salem, Oregon 97310

George E. Gerhard, Jr.
Director, In-service Training
Columbia Park Hospital and
Training Center
400 East Scenic Drive
The Dalles, Oregon 97058

Dr. R. L. Guiss
Superintendent
Damasch State Hospital
Post Office Box 38
Wilsonville, Oregon 97070

James G. Hansen, Director
Curry County Family Service Clinic
Post Office Box 1066
Gold Beach, Oregon 97444

Paul Hargis
Research Analyst
Oregon Board of Control
Salem, Oregon 97310

Garth T. Harrison, Director
Josephine County Child Guidance Clinic
205 Northwest "C" Street
Grants Pass, Oregon 97526

Dr. Fred W. Heryford
Superintendent
Wyoming State Training School
Lander, Wyoming

Glen Hitchcock
Research Analyst
Oregon Board of Control
Salem, Oregon 97310

Gladys Ingham
Administrative Officer
Alaska Psychiatric Institute
2900 Providence Avenue
Anchorage, Alaska 99504

Dr. Jay M. Jackson
Professor of Sociology
University of Oregon
Eugene, Oregon 97403

Ronald Jasperson
Department of Institutions
Post Office Box 768
Olympia, Washington 98501

Dr. N. B. Jetmalani, Director
Education and Research
Oregon State Hospital
Salem, Oregon 97310

Francis Kester, Chief
Bureau of Vital Statistics
Alaska Department of Health
and Welfare
Juneau, Alaska 99801

Glenn W. Kunsman
Acting Superintendent
Columbia Park Hospital and
Training Center
400 East Scenic Drive
The Dalles, Oregon 97058

Marian H. Larson
Director of Nursing Service
Damasch State Hospital
Post Office Box 38
Wilsonville, Oregon 97070

Fred E. Letz
Psychiatric Social Work Consultant
Community Services Section
Mental Health Division
5 State Capitol
Salem, Oregon 97310

Dr. Martin Levine
Clinical Psychologist
Multnomah County Mental Health Clinic
12240 Northeast Glisan Street
Portland, Oregon 97230

Dr. Jean Macfarlane, Director
Outpatient Department
Fairview Hospital and Training Center
2250 Strong Road, S.E.
Salem, Oregon 97310

Dr. John R. Marks
Superintendent
Idaho State School and Hospital
Nampa, Idaho 83651

Dr. Paul McCullough
Senior Clinical Psychologist
Las Vegas Mental Health Center
1000 Shadow Lane
Las Vegas, Nevada 89106

Adele McHale
Medical Records Librarian
Oregon State Hospital
Salem, Oregon 97310

Charles M. McLean
Administrator, Division of Mental
Health
Alaska Department of Health and
Welfare
Juneau, Alaska 99801

Dr. M. W. Neal, Chief
Office of Program Review
California Department of Mental
Hygiene
Sacramento, California 95814

Charles J. Nelson
Executive Secretary
Inter-Agency Committee on Mental
Retardation
5 State Capitol
Salem, Oregon 97310

Thomas F. Nugent, Director
Lane County Mental Health Clinic
Lane County Court House
Eugene, Oregon 97401

Gordon Osborne
Mental Health Consultant
Division of Mental Health
5 State Capitol
Salem, Oregon 97310

Ulissee C. Pardini, Head
Research and Data Processing
Idaho Department of Health
Boise, Idaho 83701

Dr. G. A. Parvaresh
Clinical Director
Damasch State Hospital
Post Office Box 38
Wilsonville, Oregon 97070

Lloyd E. Peterson, Director
Polk County Mental Health Clinic
Polk County Court House
Dallas, Oregon 97338

Dr. James M. Pomeroy
Superintendent
Fairview Hospital and Training
Center
2250 Strong Road, S.E.
Salem, Oregon 97310

William J. Price, Director
Yamhill County Guidance Service
Yamhill County Court House
McMinnville, Oregon 97128

Dr. Glen A. Purdom, Jr.
Director of Staff Development
Fairview Hospital and Training
Center
2250 Strong Road, S.E.
Salem, Oregon 97310

Duane Ragan
Program Consultant
Day Care for the Trainable Retarded
Mental Health Division
5 State Capitol
Salem, Oregon 97310

Dr. John P. Rollins
Chief, Psychiatric Service
Alaska Psychiatric Institute
2900 Providence Avenue
Anchorage, Alaska 99504

Charlotte Ruiz
Psychologist
Lane County Mental Health Clinic
Lane County Court House
Eugene, Oregon

Richard Runyon, Director
Clatsop County Mental Health Clinic
Fourth Floor, Columbia Hospital
Astoria, Oregon 97103

Dr. George Saslow
Professor and Chairman
Department of Psychiatry
University of Oregon Medical School
Portland, Oregon 97201

Edwin H. Shaw, Director
Coos County Family Service Clinic
Room 8, U. S. National Bank Bldg.
Coquille, Oregon 97423

Dr. E. I. Silk
Superintendent
Eastern Oregon Hospital and
Training Center
Pendleton, Oregon 97801

Robert W. Smith
Legislative Analyst
Legislative Fiscal Committee
311 State Capitol
Salem, Oregon 97310

Gertrude Stoltenberg
Director of Social Service
Oregon State Hospital
Salem, Oregon 97310

John F. Talmadge, Director
Linn County Guidance Clinic
521 Lyon Street
Albany, Oregon 97321

Aubrey D. Taylor
Research Analyst
Mental Health Division
5 State Capitol
Salem, Oregon 97310

Dr. Eugene E. Taylor, Director
Community Services Section
Mental Health Division
5 State Capitol
Salem, Oregon 97310

Wesley B. Terwilliger
Home-Family Project Director
Eastern Oregon Hospital and
Training Center
Pendleton, Oregon 97801

Dr. Ronald Tiffany
Clinical Psychologist
Mid-Columbia Mental Health Clinic
502 East Fifth Street
The Dalles, Oregon 97058

Dr. J. H. Treleaven
Clinical Director
Oregon State Hospital
Salem, Oregon 97310

Dr. George C. Turner
Chief Psychologist
Oregon State Hospital
Salem, Oregon 97310

Inez E. Veatch
Director of Nursing Service
Columbia Park Hospital and
Training Center
400 East Scenic Drive
The Dalles, Oregon 97058

Blair Whipple
Administrator
Malheur County Mental Health
Clinic
Ontario, Oregon 97914

Charles D. Whitchurch, Director
Tri-County Family Counseling
Clinic
Deschutes County Court House
Bend, Oregon 97701

Clifton M. Wignall
Research Sociologist
Colorado State Hospital
Pueblo, Colorado 81003

Cecil R. Wurster, Chief
Statistical Program Development
Section, Office of Biometry
National Institute of Mental Health
Bethesda, Maryland 20014

WICHE STAFF

Charles W. Pettus, Director
Western Conference on the Uses of
Mental Health Data

Susan E. Treece, Secretary

29:5C:367:WICHE